

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

[12] 发明专利申请公开说明书

C12N 15/82
C12N 5/04 C12N 15/52
C12N 9/00 C07K 14/415
A01H 1/00

[21] 申请号 00803955.0

[43]公开日 2002年3月20日

[11]公开号 CN 1341151A

[22]申请日 2000.1.13 [21]申请号 00803955.0

[30]优先权

[32]1999.1.15 [33]US [31]09/232,760

[32]1999.1.26 [33]US [31]09/237,479

[32]1999.2.3 [33]US [31]09/244,288

[32]1999.2.18 [33]US [31]09/252,336

[32]1999.3.30 [33]US [31]09/281,376

[86]国际申请 PCT/EP00/00246 2000.1.13

[87]国际公布 WO00/42205 英 2000.7.20

[85]进入国家阶段日期 2001.8.17

[71]申请人 辛根塔参与股份公司

地址 瑞士巴塞尔

[72]发明人 J·Z·莱文 G·J·巴德兹泽维斯克

S·L·波特勒维斯

L·维格里音格劳弗

[74]专利代理机构 中国国际贸易促进委员会专利商标事
务所

代理人 黄革生

权利要求书 38 页 说明书 56 页 附图页数 0 页

[54]发明名称 除草剂的靶基因和方法

[57]摘要

本发明涉及分离自鼠耳芥属的编码幼苗生长必需蛋白质的基因。基于这些基因对于正常生长和发育的重要性,本发明还包括应用这些蛋白质 开发新的除草剂的方法。本发明还可以用于鉴别是潜在除草剂的抑制剂的筛选分析方法中。还可以应用本发明开发除草剂耐受性植物、植物组织、植物种子和植物细胞。

I S S N 1 0 0 8 - 4 2 7 4

权利要求书

1. 分离的 DNA 分子，其含有与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列基本相似的核苷酸序列。

2. 权利要求 1 的 DNA 分子，其中该序列编码与选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列基本相似的氨基酸序列。

3. 权利要求 1 的 DNA 分子，其中该序列是选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列。

4. 权利要求 1 的 DNA 分子，其中该序列编码选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列的氨基酸序列。

5. 根据权利要求 1 的 DNA 分子，其中所述核苷酸序列是植物的核苷酸序列。

6. 权利要求 5 的 DNA 分子，其中该植物是鼠耳芥。

7. 权利要求 1 的 DNA 分子，其中该蛋白质具有选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性。

8. 氨基酸序列，其含有与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列基本相似的核苷酸序列所编码的氨基酸序列。

9. 权利要求 8 的氨基酸序列, 其含有由选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列编码的氨基酸序列。

10. 氨基酸序列, 其含有与选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列基本相似的氨基酸序列。

11. 权利要求 10 的氨基酸序列, 其中该序列是选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列。

12. 权利要求 8 的氨基酸序列, 其中该蛋白质具有选自 245、5283、2490、3963 和 4036 活性的任意一种活性。

13. 氨基酸序列, 其含有由选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列所编码的氨基酸序列的至少 20 个连续氨基酸残基。

14. 氨基酸序列, 其含有选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的氨基酸序列的至少 20 个连续氨基酸残基。

15. 含有可操作地与根据权利要求 1 的 DNA 分子连接的启动子的表达盒。

16. 含有根据权利要求 15 的表达盒的重组载体, 其中所述载体能够被稳定地转化至宿主细胞中。

17. 含有根据权利要求 15 的表达盒的宿主细胞, 其中所述核苷酸序

列能够在所述细胞中表达。

18. 根据权利要求 17 的宿主细胞，其中所述宿主细胞是真核细胞。
19. 根据权利要求 17 的宿主细胞，其中所述宿主细胞选自昆虫细胞、酵母细胞和植物细胞。
20. 根据权利要求 17 的宿主细胞，其中所述宿主细胞是原核细胞。
21. 根据权利要求 17 的宿主细胞，其中所述宿主细胞是细菌细胞。
22. 含有权利要求 19 的植物细胞的植物或种子。
23. 权利要求 22 的植物，其中所述植物对选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的抑制剂具有耐受性。
24. 制备编码具有改变活性的基因产物的核苷酸序列的方法，所述活性选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性，该方法包括
 - a) 改组权利要求 1 的核苷酸序列，
 - b) 表达获得的改组核苷酸序列，和
 - c) 与选自所述未修饰核苷酸序列基因产物的 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的活性比较，选择改变的选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的活性。
25. 权利要求 24 的方法，其中该核苷酸序列是选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列。
26. 通过权利要求 24 的方法可以获得的改组 DNA 分子。

27. 通过权利要求 24 的方法制备的重组 DNA 分子。

28. 通过权利要求 24 的方法获得的重组 DNA 分子，其中所述重组 DNA 分子编码对选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的抑制剂具有增强耐受性的基因产物。

29. 表达盒，其含有可操作地与根据权利要求 26 的核苷酸序列连接的启动子。

30. 含有根据权利要求 29 的表达盒的重组载体，其中所述载体能够被稳定地转化至宿主细胞中。

31. 含有根据权利要求 29 的表达盒的宿主细胞，其中所述核苷酸序列能够在所述细胞中表达。

32. 根据权利要求 31 的宿主细胞，其中所述宿主细胞是真核细胞。

33. 根据权利要求 31 的宿主细胞，其中所述宿主细胞选自昆虫细胞、酵母细胞和植物细胞。

34. 根据权利要求 31 的宿主细胞，其中所述宿主细胞是原核细胞。

35. 根据权利要求 31 的宿主细胞，其中所述宿主细胞是细菌细胞。

36. 含有权利要求 33 的植物细胞的植物或种子。

37. 权利要求 36 的植物，其中所述植物对选自抑制 245、5283、2490、3963 和 4036 活性的抑制剂具有耐受性。

38. 筛选与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列所编码的蛋白质相互作用的化合物的方法, 包括:

- a) 表达含有选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列, 或与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列基本相似的序列的 DNA 分子, 以产生相应的蛋白质,
- b) 测试怀疑能够与步骤 (a) 中表达的蛋白质相互作用的化合物, 和
- c) 选择在步骤 (b) 中与所述蛋白质相互作用的化合物。

39. 鉴定选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的抑制剂的方法, 包括:

- a) 将含有选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列的核苷酸序列并具有选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的 DNA 分子, 或与之基本相似的核苷酸序列, 或它们的同系物, 导入植物细胞, 以致所述序列能够以高于野生型表达水平的水平进行功能性表达,
- b) 在有利于抑制发生的条件下将所述植物细胞与待测试的化合物合并, 以检测该化合物对选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的抑制能力,
- c) 在步骤 b) 的条件下测量植物细胞的生长, 和
- d) 在相同条件下, 比较所述植物细胞的生长和具有选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的未改变活性的植物细胞的生长, 和
- e) 选择在步骤 d) 中抑制植物细胞生长的所述化合物。

40. 根据权利要求 39 的方法可鉴定的具有除草剂活性的化合物。

41. 鉴定具有除草剂活性的化合物的方法，包括：

- a) 在有利于相互作用的条件下，使权利要求 8 的蛋白质与待测化合物混合，以测试该化合物与所述蛋白质相互作用的能力，
- b) 选择在步骤 (a) 中鉴定的能够与所述蛋白质相互作用的化合物。
- c) 给植物施用步骤 (b) 中鉴定的化合物，以测试其除草剂活性，和
- d) 选择具有除草剂活性的化合物。

42. 根据权利要求 41 的方法可鉴定的具有除草剂活性的化合物。

43. 抑制植物生长的方法，包括以足以抑制所述植物生长的量给所述植物施用抑制权利要求 8 的氨基酸序列的活性的化合物。

44. 权利要求 41 的方法，其中该化合物是根据权利要求 39 的方法可鉴定的具有除草剂活性的化合物。

45. 改良农作物的方法，包括给选自权利要求 23 的植物或种子和权利要求 37 的植物或种子的除草剂耐受性植物或种子，施用根据选自权利要求 38 的方法、权利要求 39 的方法和权利要求 41 的方法的方法可鉴定的具有除草剂活性的化合物，施用量为抑制不需要的植物的生长但不显著抑制该除草剂耐受性植物或种子的生长的量。

46. DNA 分子，其含有与选自 SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20、SEQ ID NO:21、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 的任意一个序列基本相似的核苷酸序列。

说明书

除草剂的靶基因和方法

本发明涉及从鼠耳芥属 (*Arabidopsis*) 分离的编码幼苗生长所必需的蛋白质的基因。基于该基因对于正常生长和发育的重要性, 本发明还包括将这些蛋白质用作除草剂靶标的方法。本发明还可以用于鉴定是潜在除草剂的抑制剂的筛选方法。本发明还可以应用于开发除草剂耐受性植物、植物组织、植物种子和植物细胞。

应用除草剂控制农作物田间不需要的植物例如杂草几乎已成为一种全球性手段。每年除草剂的销售超过 150 亿美元。尽管有这种广泛应用, 但杂草的控制对于农民而言仍然是一个重要而耗资很大的问题。

除草剂的有效应用需要合理的安排。例如, 使用的时间和方法以及杂草植物的发育阶段对于使用除草剂获得良好的杂草控制是关键。由于有多种杂草种类可以抵抗除草剂, 所以制备有效的新除草剂变得越来越重要。目前可以采用以重组 DNA 技术进行的大批量筛选方法发现新的除草剂。植物的生长和发育所必需的代谢酶可以通过标准分子生物学技术来重组制备, 并作为除草剂的靶标应用于新的该酶活性抑制剂的筛选中。通过此筛选发现的新抑制剂然后可以用作除草剂控制不需要的植物。

不幸的是, 表现出更大的效力、更广的杂草范围以及能够更快地在土壤中降解的除草剂可能也具有更大的危害农作物的植物毒性。该问题的一个解决方法是开发抵抗或耐受除草剂的农作物。耐受除草剂的农作物杂种或品体使得可以应用这些除草剂来杀灭杂草而又不伴随对农作物造成伤害的危险。耐受性的开发能够允许除草剂应用于原先由于对该除草剂敏感而不能或限制性地 (例如出苗前使用) 使用该除草剂的农作物上。例如, Anderson 等的美国专利 4, 761, 373 针对多种咪唑啉酮或磺胺类除草剂的抗性植物。一种改变的乙酰羟酸合成酶 (AHAS) 赋予了该抗性。Goodman 等的美国专利 4, 975, 374 涉及到含有编码突变的谷氨酰胺合

成酶 (GS) 的基因的植物细胞和植物, 这些植物细胞和植物抵抗已知抑制 GS 的除草剂例如膦丝菌素和甲硫氨酸磺基肟 (methionine sulfoximine) 的抑制。Bedbrook 等的美国专利 5, 013, 659 针对表达突变的乙酰乳酸合成酶的植物, 该酶使得该植物能够抵抗磺酰脲除草剂的抑制。Somers 等的美国专利 5, 162, 602 公开了耐受环己烷二酮和芳氧基苯氧基丙酸除草剂抑制的植物。该耐受性由改变的乙酰辅酶 A 羧化酶 (ACCase) 赋予。

尽管有以上描述的进展, 但对于不需要或有害的植物生长 (例如杂草) 仍存顽固性和亟待解决的问题。而且, 随着人口的持续生长, 将出现日愈增加的食品短缺。因此, 对于寻找新的有效而经济的除草剂, 存在一种长期感受到但未满足的需求。

为了清楚起见, 将本说明书中使用到的某些术语定义和介绍如下:

嵌合的: 用以指 DNA 序列, 例如载体或基因, 是由一种以上不同来源的 DNA 序列组成的其中这些不同来源的 DNA 序列通过重组 DNA 技术融合在一起, 形成一种天然不存在的, 尤其是在待转化的植物中不存在的 DNA 序列。

辅因子: 酶催化反应中所必需的天然反应物, 例如有机分子或金属离子。辅因子例如 NAD(P)、核黄素 (包括 FAD 和 FMN)、叶酸、钼蝶呤 (molybdopterin)、硫胺素、生物素、硫辛酸、泛酸和辅酶 A、S-腺苷甲硫氨酸、吡哆醛磷酸、辅酶 Q、甲基萘醌类。任选地, 辅因子可以再生和重利用。

DNA 改组: DNA 改组是一种快速容易并有效地在 DNA 分子中引入, 优选随机地引入突变或重排, 或在两种或多种 DNA 分子之间产生, 优选随机地产生 DNA 序列交换的方法。从 DNA 改组获得的 DNA 分子是一种改组的 DNA 分子, 该分子是来源于至少一种模板 DNA 分子的天然所不存在的 DNA 分子。该改组 DNA 编码就模板 DNA 所编码的酶而言修饰的酶, 而且相对于模板 DNA 所编码的酶而言优选具有改变的生物学活性。

酶活性: 在本文中指酶催化底物转化成产物的能力。酶的底物包含该酶的天然底物, 还包含也能被该酶转化成产物或产物类似物的该天然

底物的类似物。酶活性的测量通过在一个确定的时间段之后测定反应中的产物量，或在一个确定的时间段之后测定反应混合物中剩余的底物量来进行。还可以通过在一个确定的时间段之后测定反应混合物中剩余的未使用辅因子的量，或通过在一个确定的时间段之后测定反应混合物中用掉的辅因子的量来测量酶活性。还可以通过在一个确定的时间段之后测定反应混合物中剩余的自由能供体或富含能量分子（例如 ATP、磷酸烯醇丙酮酸、乙酰磷酸或磷酸肌酸）的量，或通过在一个确定的时间段之后测定反应混合物中已用自由能供体或富含能量分子（例如 ADP、丙酮酸、乙酸或肌酸）的量来测量酶活性。

表达：是指植物中内源性基因或转基因的转录和/或翻译。例如对于反义结构，表达可以仅指该反义 DNA 的转录。

基因：是指编码序列和相关的调节序列，其中该编码序列转录成 RNA 例如 mRNA、rRNA、tRNA、snRNA、有义 RNA 或反义 RNA。调节序列的例子是启动子序列、5'和 3'非翻译序列和终止序列。还可以存在的元件是例如内含子。

除草剂：用以杀死或抑制植物、植物细胞、植物种子或植物组织生长的化学制品。

异源 DNA 序列：天然情况下与所导入的宿主细胞不相关的 DNA 序列，包括天然存在 DNA 序列的非天然存在的多拷贝，以及其中本身同源的 DNA 序列可操作地与非天然的序列连接的遗传结构。

同源 DNA 序列：与其导入的宿主细胞天然相关的 DNA 序列。

抑制剂：通过例如失活对于植物的生长或生存所必需的生物合成酶等蛋白质的酶活性、受体、信号转导蛋白质、结构基因产物或运输蛋白，造成植物异常生长的化学制品。在本发明的上下文中，抑制剂是指改变植物的 245 基因、5283 基因、2490 基因、3963 基因或 4036 基因所编码的酶活性的化学制品。更一般地，抑制剂通过与该 245 基因、5283 基因、2490 基因、3963 基因或 4036 基因所编码的基因产物相互作用造成宿主细胞的异常生长。

等基因的：除了可能由于存在或不存在异源 DNA 序列而不同外，遗

传上一致的植物。

分离的：在本发明的上下文中，分离的 DNA 分子或分离的酶是指通过人为操作使其脱离天然环境而存在并因此不是自然产物的 DNA 分子或酶。分离的 DNA 分子或酶可以以纯化形式存在，或可以存在于非天然环境中例如存在于转基因宿主细胞中。

标记基因：编码可筛选或可选择的性状的基因。

成熟蛋白质：正常定向到细胞器例如叶绿体，并且已从其中除去了转运肽的蛋白质。

最小启动子：在没有上游激活时无活性或具有大大降低的启动子活性的启动子元件，尤其是 TATA 元件。当存在适合的转录因子时，该最小启动子发挥功能允许转录。

修饰的酶活性：与天然存在于植物中的酶活性（即在没有人对地对这些活性进行直接或间接的操作的情况下天然存在的酶活性）不同的酶活性，该修饰的酶活性耐受抑制天然存在的酶活性的抑制剂。

植物：是指任何植物，尤其是种子植物。

植物细胞：植物的结构和生理单位，包含原生质和细胞壁。植物细胞可以是分离的单个细胞或培养细胞形式、或作为更高级的组织单位例如植物组织或植物器官等的部分。

植物材料：是指植物的叶、茎、根、花或花的部分、果实、花粉、花粉管、胚珠、胚囊、卵细胞、合子、胚、种子、插条、细胞或组织培养物、或任何其它的植物部分或植物产物。

前蛋白 (pre-protein)：正常定向到细胞器例如叶绿体，并且仍含有其转运肽的蛋白质。

重组 DNA 分子：采用重组 DNA 技术连接在一起的 DNA 序列的组合。

选择标记基因：其表达并不赋予转化细胞选择优势，但其表达使得该转化细胞表型上不同于非转化细胞的基因。

显著增加：大于测量技术固有的误差范围的酶活性增加，优选在有抑制剂时比野生型酶的活性增加大约 2 倍或更多，更优选增加大约 5 倍或更多，最优选增加大约 10 倍或更多。

显著降低: 是指酶反应产物量的降低超过测量技术固有的误差范围。优选在没有抑制剂时比野生型酶的活性降低大约 2 倍或更多, 更优选降低大约 5 倍或更多, 最优选降低大约 10 倍或更多。

术语“基本相似”当在本文中用于核苷酸序列时, 以其最广泛的含义, 是指相应于参考序列的核苷酸序列, 其中该相应序列所编码的多肽具有与该参考核苷酸序列编码的多肽基本相似的结构和功能, 例如仅在不影响多肽功能的氨基酸中发生改变。期望的是, 该基本相似的核苷酸序列编码由该参考核苷酸序列编码的多肽。术语“基本相似”特别旨在包括其中序列已被修饰以优化其在特定细胞中的表达的核苷酸序列。该基本相似的核苷酸序列与参考核苷酸序列之间的一致性百分数希望是至少 65%、更希望是至少 75%、优选至少 85%、更优选至少 90%、甚至更优选至少 95%、甚至更优选至少 99%。序列比较采用 Smith-Waterman 序列比对算法 (见例如 Waterman, M. S., 计算生物学入门: 图谱、序列和基因组 (Introduction to Computational Biology: Maps, Sequences and genomes), Chapman & Hall, 伦敦: 1995, ISBN 0-412-99391-0) 进行。使用 1.16 版 localS 程序, 参数如下: 匹配: 1, 错配罚分: 0.33, 开始断缺区罚分: 2, 延长的断缺区罚分: 2。与参考核苷酸序列“基本相似”的核苷酸序列可以与该参考核苷酸序列在如下条件下杂交: 7% 十二烷基磺酸钠 (SDS)、0.5M NaPO₄、1mM EDTA 中 50℃ 杂交并在 2× SSC、0.1% SDS 中 50℃ 进行洗涤, 更期望的是在 7% 十二烷基磺酸钠 (SDS)、0.5M NaPO₄、1mM EDTA 中 50℃ 杂交并在 1× SSC、0.1% SDS 中 50℃ 进行洗涤, 甚至更期望的是在 7% 十二烷基磺酸钠 (SDS)、0.5M NaPO₄、1mM EDTA 中 50℃ 杂交并在 0.5× SSC、0.1% SDS 中 50℃ 进行洗涤, 优选在 7% 十二烷基磺酸钠 (SDS)、0.5M NaPO₄、1mM EDTA 中 50℃ 杂交并在 0.1× SSC、0.1% SDS 中 50℃ 进行洗涤, 更优选在 7% 十二烷基磺酸钠 (SDS)、0.5M NaPO₄、1mM EDTA 中 50℃ 杂交并在 0.1× SSC、0.1% SDS 中 65℃ 进行洗涤。本文所用术语“245 基因”、“5283 基因”、“2490 基因”、“3963 基因”、或“4036 基因”是指分别含有 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9, 或分别含有与 SEQ ID NO:1、SEQ

ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 基本相似的核苷酸序列的 DNA 分子。245 基因、5283 基因、2490 基因、3963 基因或 4036 基因的同系物分别包括编码与 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10 有至少 30% 一致性的氨基酸序列的核苷酸序列，一致性的测量采用以下描述的参数进行，其中由这些同系物编码的氨基酸序列分别具有 245、5283、2490、3963 或 4036 蛋白质的生物学活性。

术语“基本相似”，当在本文中用于蛋白质时，是指相应于参考蛋白质的蛋白质，其中该蛋白质具有与该参考蛋白质基本相同的结构和功能，例如仅在不影响该多肽功能的氨基酸序列中有改变。当用于蛋白质或氨基酸序列时，采用 BLAST 分析的默认参数值，基本相似的蛋白质或氨基酸序列和参考蛋白质或氨基酸序列之间的一致性百分数期望是至少 65%、更期望至少 75%、优选至少 85%、更优选至少 90%、甚至更优选至少 95%、甚至更优选至少 99%。本文所用术语“245 蛋白质”、“5283 蛋白质”、“2490 蛋白质”、“3963 蛋白质”、或“4036 蛋白质”分别是指由含有与 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 基本相似的核苷酸序列的 DNA 分子编码的氨基酸序列。245 蛋白质、5283 蛋白质、2490 蛋白质、3963 蛋白质或 4036 蛋白质的同系物分别是与 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10 有至少 30% 一致性的氨基酸序列，一致性的测量采用以下描述的参数进行，其中这些同系物分别具有 245、5283、2490、3963 或 4036 蛋白质的生物学活性。

本领域技术人员还熟悉确定“基本相似的”和参考的核苷酸序列、或蛋白质或氨基酸序列之间一致性百分数的其它分析工具，例如 GAP 分析。因此，在本发明中，“基本相似”也采用 Wisconsin 大学 GCG，GAP 的 SEQWEB 应用程序以及默认的 GAP 分析参数来确定，该 GAP 分析基于 Needleman 和 Wunsch 的算法 (Needleman 和 Wunsch (1970)，分子生物学杂志 (J. Mol. Biol.) 48:443-453)。

因此，对于“245 基因”并采用上述 GAP 分析时，“基本相似”是

指编码与 SEQ ID NO:2 有至少 47% 一致性、更优选至少 60% 一致性、甚至更优选至少 75% 一致性、甚至更优选至少 85% 一致性、甚至更优选至少 95% 一致性、甚至更优选至少 99% 一致性的蛋白质的核苷酸序列。

对于“5283 基因”并采用上述 GAP 分析时，“基本相似”是指编码与 SEQ ID NO:4 有至少 74% 一致性、更优选至少 80% 一致性、甚至更优选至少 85% 一致性、甚至更优选至少 90% 一致性、甚至更优选至少 95% 一致性、甚至更优选至少 99% 一致性的蛋白质的核苷酸序列。而且，“基本相似”还优选指与 SEQ ID NO:3 有至少 80% 一致性、更优选至少 90% 一致性、甚至更优选 95% 一致性、甚至更优选至少 99% 一致性的核苷酸序列，其中所述核苷酸序列比较采用上述 GAP 分析进行。

对于“2490 基因”并采用上述 GAP 分析时，“基本相似”是指编码与 SEQ ID NO:6 有至少 82% 一致性、更优选至少 85% 一致性、甚至更优选至少 90% 一致性、甚至更优选至少 95% 一致性、甚至更优选至少 99% 一致性的蛋白质的核苷酸序列。而且，“基本相似”还优选指与 SEQ ID NO:5 有至少 87% 一致性、更优选至少 90% 一致性、甚至更优选 95% 一致性、甚至更优选至少 99% 一致性的核苷酸序列，其中所述核苷酸序列比较采用上述 GAP 分析进行。

对于“3963 基因”并采用上述 GAP 分析时，“基本相似”是指编码与 SEQ ID NO:8 有至少 40% 一致性、更优选至少 60% 一致性、更优选至少 80% 一致性、甚至更优选至少 90% 一致性、甚至更优选至少 95% 一致性、甚至更优选至少 99% 一致性的蛋白质的核苷酸序列。而且，“基本相似”还优选指与 SEQ ID NO:7 有至少 49% 一致性、更优选至少 60% 一致性、甚至更优选至少 80% 一致性、更优选至少 90% 一致性、更优选至少 95% 一致性、甚至更优选至少 99% 一致性的核苷酸序列，其中所述核苷酸序列比较采用上述 GAP 分析进行。

对于“4036 基因”并采用上述 GAP 分析时，“基本相似”是指编码与 SEQ ID NO:10 有至少 67% 一致性、更优选至少 80% 一致性、更优选至少 85% 一致性、甚至更优选至少 90% 一致性、甚至更优选至少 95% 一致性、甚至更优选至少 99% 一致性的蛋白质的核苷酸序列。

而且，采用如上所述的 GAP 分析，“245 基因的同系物”包括编码与 SEQ ID NO:2 有至少 24% 一致性、更优选至少 30% 一致性、甚至更优选至少 40% 一致性、甚至更优选至少 45% 一致性、甚至更优选至少 55% 一致性、甚至更优选至少 65% 一致性、甚至更优选至少 75% 一致性的氨基酸序列的核苷酸序列，其中该同系物所编码的氨基酸序列具有 245 蛋白质的生物学活性。

而且，采用如上所述的 GAP 分析，“5283 基因的同系物”包括编码与 SEQ ID NO:4 有至少 23% 一致性、更优选至少 40% 一致性、甚至更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 74% 一致性的氨基酸序列的核苷酸序列，其中该同系物所编码的氨基酸序列具有 5283 蛋白质的生物学活性。

而且，采用如上所述的 GAP 分析，“2490 基因的同系物”包括编码与 SEQ ID NO:6 有至少 30% 一致性、更优选至少 30% 一致性、甚至更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 80% 一致性的氨基酸序列的核苷酸序列，其中该同系物所编码的氨基酸序列具有 2490 蛋白质的生物学活性。

而且，采用如上所述的 GAP 分析，“3963 基因的同系物”包括编码与 SEQ ID NO:8 有至少 34% 一致性、更优选至少 40% 一致性、甚至更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 75% 一致性的氨基酸序列的核苷酸序列，其中该同系物所编码的氨基酸序列具有 3963 蛋白质的生物学活性。

而且，采用如上所述的 GAP 分析，“4036 基因的同系物”包括编码与 SEQ ID NO:10 有至少 44% 一致性、更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 75% 一致性的氨基酸序列的核苷酸序列，其中该同系物所编码的氨基酸序列具有 4036 蛋白质的生物学活性。

当采用上述关于蛋白质或氨基酸序列的 GAP 分析时，对于“245 基因”，“基本相似”的蛋白质或氨基酸序列与参考蛋白质或氨基酸序列（在此为 SEQ ID NO:2）之间的一致性百分数为至少 47%、更优选至少 60%、

甚至更优选至少 75%、甚至更优选至少 85%、甚至更优选至少 95%、甚至更优选至少 99%。“245 蛋白质的同系物”包括与 SEQ ID NO:2 有至少 24% 一致性、更优选至少 30% 一致性、甚至更优选至少 40% 一致性、甚至更优选至少 45% 一致性、甚至更优选至少 55% 一致性、甚至更优选至少 65% 一致性、甚至更优选至少 75% 一致性的氨基酸序列，其中该 245 蛋白质的同系物具有 245 蛋白质的生物学活性。

对于“5283 基因”，当采用上述 GAP 分析时，基本相似的蛋白质或氨基酸序列与参考蛋白质或氨基酸序列（在此为 SEQ ID NO:4）之间的一致性百分数为至少 74%、更优选至少 80%、甚至更优选至少 85%、甚至更优选至少 90%、甚至更优选至少 95%、甚至更优选至少 99%。“5283 蛋白质的同系物”包括与 SEQ ID NO:4 有至少 23% 一致性、更优选至少 40% 一致性、甚至更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 74% 一致性的氨基酸序列，其中该 5283 蛋白质的同系物具有 5283 蛋白质的生物学活性。

对于“2490 基因”，当采用上述 GAP 分析时，基本相似的蛋白质或氨基酸序列与参考蛋白质或氨基酸序列（在此为 SEQ ID NO:6）之间的一致性百分数为至少 82%、更优选至少 85%、更优选至少 90%、甚至更优选至少 95%、甚至更优选至少 99%。“2490 蛋白质的同系物”包括与 SEQ ID NO:6 有至少 30% 一致性、更优选至少 30% 一致性、甚至更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 80% 一致性的氨基酸序列，其中该 2490 蛋白质的同系物具有 2490 蛋白质的生物学活性。

对于“3963 基因”，当采用上述 GAP 分析时，“基本相似”的蛋白质或氨基酸序列与参考蛋白质或氨基酸序列（在此为 SEQ ID NO:8）之间的一致性百分数为至少 40%、更优选至少 60%、更优选至少 80%、甚至更优选至少 90%、甚至更优选至少 95%、甚至更优选至少 99%。“3963 蛋白质的同系物”包括与 SEQ ID NO:8 有至少 34% 一致性、更优选至少 40% 一致性、甚至更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 75% 一致性的氨基酸序列，其中该 3963 蛋白质的同系物

具有 3963 蛋白质的生物学活性。

对于“4036 基因”，当采用上述 GAP 分析时，“基本相似”的蛋白质或氨基酸序列与参考蛋白质或氨基酸序列（在此为 SEQ ID NO:10）之间的一致性百分数为至少 67%、更优选至少 80%、更优选至少 85%、甚至更优选至少 90%、甚至更优选至少 95%、甚至更优选至少 99%。“4036 蛋白质的同系物”包括与 SEQ ID NO:10 有至少 44% 一致性、更优选至少 50% 一致性、甚至更优选至少 60% 一致性、甚至更优选至少 75% 一致性的氨基酸序列，其中该 4036 蛋白质的同系物具有 4036 蛋白质的生物学活性。

底物：底物是在酶天然执行其功能的生物化学途径中被该酶天然识别并转化成产物的分子，或是在类似于天然存在的反应的酶反应中被该酶识别并转化成产物的该分子的修饰版本。

耐受性：当暴露于足以抑制天然未修饰植物正常生长或功能的量的抑制剂或除草剂时，保持基本正常的生长或功能的能力。

转化：将异源 DNA 引入细胞、组织或植物的程序。转化的细胞、组织或植物应理解为不仅包括转化程序的终产物，也包括其转基因后代。

转基因的：用优选含有可操作地与目的 DNA 序列连接的适合启动子的重组 DNA 分子稳定转化的。

SEQ ID NO:1 鼠耳芥属 245 基因的 cDNA 序列

SEQ ID NO:2 SEQ ID NO:1 所示鼠耳芥属 245 DNA 序列所编码的氨基酸序列

SEQ ID NO:3 鼠耳芥属 5283 基因的 cDNA 序列

SEQ ID NO:4 SEQ ID NO:3 所示鼠耳芥属 5283 DNA 序列所编码的氨基酸序列

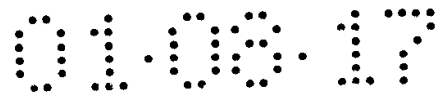
SEQ ID NO:5 鼠耳芥属 2490 基因的 cDNA 序列

SEQ ID NO:6 SEQ ID NO:5 所示鼠耳芥属 2490 DNA 序列所编码的氨基酸序列

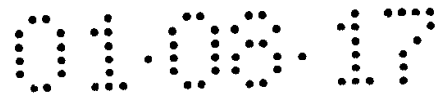
SEQ ID NO:7 鼠耳芥属 3963 基因的 cDNA 序列

- SEQ ID NO:8 SEQ ID NO:7 所示鼠耳芥属 3963 DNA 序列所编码的氨基酸序列
- SEQ ID NO:9 鼠耳芥属 4036 基因的 cDNA 序列
- SEQ ID NO:10 SEQ ID NO:9 所示鼠耳芥属 4036 DNA 序列所编码的氨基酸序列
- SEQ ID NO:11 寡核苷酸 SLP346for
- SEQ ID NO:12 鼠耳芥属 245 基因的部分基因组序列
- SEQ ID NO:13 鼠耳芥属 245 基因 cDNA 序列的 3'UTR
- SEQ ID NO:14 鼠耳芥属 5283 基因的基因组序列
- SEQ ID NO:15 寡核苷酸 SLP328
- SEQ ID NO:16 寡核苷酸 LW60
- SEQ ID NO:17 鼠耳芥属 5283 基因 cDNA 序列的 5'UTR
- SEQ ID NO:18 鼠耳芥属 5283 基因 cDNA 序列的 3'UTR
- SEQ ID NO:19 鼠耳芥属 2490 基因的基因组序列
- SEQ ID NO:20 鼠耳芥属 2490 基因 cDNA 的 5'UTR
- SEQ ID NO:21 鼠耳芥属 2490 基因 cDNA 序列的 3'UTR
- SEQ ID NO:22 寡核苷酸 SLP369
- SEQ ID NO:23 寡核苷酸 SLP370
- SEQ ID NO:24 鼠耳芥属 3963 基因的基因组序列
- SEQ ID NO:25 寡核苷酸 - 21
- SEQ ID NO:26 鼠耳芥属 3963 基因 cDNA 序列的 3'UTR
- SEQ ID NO:27 鼠耳芥属 4036 基因的基因组序列
- SEQ ID NO:28 鼠耳芥属 4036 基因的 cDNA 编码序列, 包括来自栽培品种 Landsberg 的 cDNA 和 Columbia 的基因组序列之间的差异。
- SEQ ID NO:29 SEQ ID NO:28 所示鼠耳芥属 4036 DNA 序列所编码的氨基酸序列

本发明包括分离的 DNA 分子, 其含有与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列基本



相似的核苷酸序列。优选根据本发明的 DNA 分子，其中该序列编码与选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列基本相似的氨基酸序列。进一步优选根据本发明的 DNA 分子，其中该序列是选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列。进一步优选根据本发明的 DNA 分子，其中该序列编码选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列的氨基酸序列。进一步优选根据本发明的 DNA 分子，其中所述核苷酸序列是植物的核苷酸序列。更优选根据本发明的 DNA 分子，其中该植物是鼠耳芥 (*Arabidopsis thaliana*)。进一步优选根据本发明的 DNA 分子，其中该蛋白质具有选自 245 活性、5283 活性、2490 活性、396 活性和 4036 活性的任意一种活性。本发明还包括氨基酸序列，其含有与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列基本相似的核苷酸序列所编码的氨基酸序列。优选根据本发明的氨基酸序列，其含有选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列所编码的氨基酸序列。本发明的再一个目标是含有与选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列基本相似的氨基酸序列的氨基酸序列。优选根据本发明的氨基酸序列，其中该序列是选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的任意一个序列。进一步优选根据本发明的 DNA 序列，其中该蛋白质具有选自 245、5283、2490、3963 和 4036 活性的任意一种活性。本发明包括氨基酸序列，其含有选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列所编码的氨基酸序列的至少 20 个连续氨基酸残基。本发明还包括含有选自 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 和 SEQ ID NO:10 的氨基酸序列的至少 20 个连续氨基酸残基的氨基酸序列。本发明的一个目标是含有可操作地与本发明 DNA 分子连接的启动子的表达盒。本发明还包括含有根据本发明的表达盒的重组载体，其中所述载体能够被稳定地转化至宿主细胞中。本发明还包括含有根据本发明



的表达盒的宿主细胞，其中所述核苷酸序列能够在所述细胞中表达。优选根据本发明的宿主细胞，其中所述宿主细胞是真核细胞。更优选根据本发明的宿主细胞，其中所述宿主细胞选自昆虫细胞、酵母细胞和植物细胞。而且更优选根据本发明的宿主细胞，其中所述宿主细胞是原核细胞。更优选根据本发明的宿主细胞，其中所述宿主细胞是细菌细胞。本发明包括含有根据本发明的植物细胞的植物或种子。优选根据本发明的植物，其中所述植物耐受选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的抑制剂。

本发明还包括如下方法，该方法包括获得含有异源 DNA 分子的宿主细胞，其中所述异源 DNA 分子编码具有 245 活性、5283 活性、2490 活性、3963 活性或 4036 活性的蛋白质；和在所述宿主细胞中表达所述蛋白质。优选该宿主细胞是细菌细胞、酵母细胞或昆虫细胞。

本发明还包括制备编码具有改变活性的基因产物的核苷酸序列的方法，其中所述活性选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性，该方法包括：

- a) 改组权利要求 1 的核苷酸序列，
- b) 表达所获得的改组核苷酸序列，和
- c) 选择与选自所述未修饰核苷酸序列的基因产物的 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性相比，改变的选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的活性。

优选根据本发明的方法，其中该核苷酸序列是选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列。本发明包括通过根据本发明的该方法可以获得的改组 DNA 分子。本发明包括通过根据本发明的该方法制备的改组 DNA 分子。本发明还包括根据本发明的该方法获得的改组 DNA 分子，其中所述改组 DNA 分子编码的基因产物对选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的抑制剂具有增强的耐受性。本发明的再一个目标是含有可操作地与本发明核苷酸序列连接的启动子的表达盒。本发明还包括含有根据本发明的表达盒的重组载体，其中所述载体能够被稳定地转化

至宿主细胞中。本发明的再一个目标是含有根据本发明的表达盒的宿主细胞，其中所述核苷酸序列能够在所述细胞中表达。优选根据本发明的宿主细胞，其中所述宿主细胞是真核细胞。还优选根据本发明的宿主细胞，其中所述宿主细胞选自昆虫细胞、酵母细胞和植物细胞。还优选根据本发明的宿主细胞，其中所述宿主细胞是原核细胞。还优选根据本发明的宿主细胞，其中所述宿主细胞是细菌细胞。本发明的一个目标是含有根据本发明的植物细胞的植物或种子。优选根据本发明的植物，其中所述植物耐受选自抑制 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的抑制剂。本发明还包括筛选与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列所编码的蛋白质相互作用的化合物的方法，包括：

a) 表达分别含有选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列，或含有与选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列基本相似的序列的 DNA 分子，以产生相应蛋白质，

b) 对推测能够与步骤 (a) 中表达的蛋白质相互作用的化合物进行测试，和

c) 选择在步骤 (b) 中与蛋白质相互作用的化合物。

本发明的另一个目标是鉴定选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的抑制剂的方法，包括：

(a) 向植物细胞中分别引入含有选自 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 和 SEQ ID NO:9 的任意一个序列的核苷酸序列，并具有选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的 DNA 分子，或与其基本相似的核苷酸序列，或它们的同系物，以致所述序列能够以高于野生型表达水平的水平进行功能性表达，

(b) 在有利于抑制的条件下将所述植物细胞与待测试化合物合并，以检测该化合物抑制选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的任意一种活性的能力，

(c) 在步骤(b)的条件下测量植物的生长, 和

(d) 在相同条件下, 将所述植物细胞的生长与具有选自 245 活性、5283 活性、2490 活性、3963 活性和 4036 活性的未改变活性的植物细胞之生长进行比较, 和

(e) 选择在步骤(d)中抑制植物细胞生长的所述化合物。

本发明包括根据本发明的方法能够鉴别的具有除草剂活性的化合物。本发明还包括鉴定具有除草剂活性的化合物的方法, 包括:

(a) 在有利于相互作用的条件下将根据本发明的蛋白质与待测试的化合物合并, 以测试该化合物与所述蛋白质相互作用的能力,

(b) 选择在步骤(a)中鉴定的能够与所述蛋白质相互作用的化合物,

(c) 给植物施用步骤(b)中所鉴定的化合物, 以检测除草剂活性, 和

(d) 选择具有除草剂活性的化合物。

本发明包括根据本发明的方法能够鉴定的具有除草剂活性的化合物。本发明的再一个目标是用于抑制植物生长的方法, 包括给所述植物施用抑制本发明氨基酸序列的活性的化合物, 施用量为足以抑制所述植物生长的量。

优选根据本发明的方法, 其中该化合物是根据本发明的方法能够鉴定的具有除草剂活性的化合物。本发明包括改良农作物的方法, 包括给根据本发明的除草剂耐受性植物或种子施用可根据本发明方法鉴定的具有除草剂活性的化合物, 施用量是对不需要的植物产生抑制但又不显著抑制该除草剂耐受性植物或种子的生长的量。本发明的一个目标是含有与选自下组的任意一个序列基本相似的核苷酸序列的 DNA 分子: SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20、SEQ ID NO:21、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29。

本发明的一个目标是提供一种鉴定新除草剂的有效而有利的办法。

本发明的一个特点是鉴定了表现出与肽释放因子 2 (Craigen 等 (1985) 美国国家科学院院刊 (Proc. Natl. Acad. Sci.), 82: 3616 - 3620; Craigen 和 Caskey (1987) Biochimie 69:1031-1041; Ito 等 (1998) 美国国家科学院院刊, 95: 8165 - 8169) 具有序列相似性的鼠耳芥属基因, 在本文中称作 245 基因。本发明的另一个特点是发现该 245 基因是幼苗生长和发育所必需的。本发明的一个优点在于, 该新发现的必需基因包含一种新的除草剂作用方式, 其使得本领域的技术人员能够容易而快速地鉴定出新的除草剂。

本发明的另一个特点是鉴定了表现出与下列基因具有序列相似性的鼠耳芥属基因, 在本文中称作 5283 基因: 粟酒裂殖酵母 (*Schizosaccharomyces pombe*) 的一个未定性基因; 酿酒酵母 (*Saccharomyces cerevisiae*) 中编码 mRNA 前体拼接所必需的一个因子的 PRP31 基因 (Weidenhammer 等 (1996) 核酸研究 (Nucleic Acids Res.) 24:1164-1170; Weidenhammer 等 (1997) 分子细胞生物学 (Mol. Cell Biol.), 17:3580-3585); *Pisum sativum* 中编码支架结构附着区 (SAR) DNA 结合蛋白的 SARBP - 1 和 SARBP - 2 (Rzepecki 等 (1995) Acta Biochim. Pol., 41:75-81); 酿酒酵母中编码能消除 IKB 之生长抑制性作用的蛋白质的 SIK1 基因 (Morin 等 (1995) 细胞生长和分化 (Cell Growth & Differentiation), 6: 789 - 798)。该 SIK1 基因产物又称作 Nop56, 其表现出是一种必需的核仁蛋白质 (Gautier 等 (1997), 分子细胞生物学, 17: 7088 - 7098)。本发明的另一个特点在于, 发现该 5283 基因是幼苗生长和发育所必需的。本发明的一个优点是, 该新发现的必需基因包含一种新的除草剂作用方式, 其使得本领域的技术人员能够容易而快速地鉴定出新的除草剂。

本发明的再一个特点是, 鉴定了在鼠耳芥属中编码与叶绿体被膜蛋白具有序列相似性的蛋白质的基因, 本文中称作 2490 基因 (Ko 等 (1995) 生物化学杂志 (The Journal of Biological Chem.) 270: 28601-28608; Wu 等 (1994) 生物化学杂志 269:32264-32271; Pang 等 (1997) 生物化学杂志 272:25623-23627)。本发明的另一个特点在于, 发现该 2490 基因

是幼苗生长和发育所必需的。本发明的一个优点是，该新发现的必需基因包含一种新的除草剂作用方式，其使得本领域的技术人员能够容易而快速地鉴定出新的除草剂。

本发明的再一个特点是鉴定了在鼠耳芥属中编码与许多 DNA 修复蛋白质具有序列相似性的蛋白质的基因，在本文中称作 3963 基因，其中所述 DNA 修复蛋白质包括 *Schizosaccharomyces pombe* 的 Rad32p (Genbank 索取号 Q09683); 人 (*Homo sapiens*) 的 hMre11 (Genbank 索取号 U37359); 和酿酒酵母的 Mre11p (Genbank 索取号 U60829) (Johzuka 和 Ogawa (1995) 遗传学 (Genetics), 139: 1521-1532; Paull 和 Gellert (1998) 分子细胞 (Molecular Cell) 1:969-979)。本发明的另一个特点在于，发现该 3963 基因是幼苗生长和发育所必需的。本发明的一个优点是，该新发现的必需基因包含一种新的除草剂作用方式，其使得本领域的技术人员能够容易而快速地鉴定出新的除草剂。

本发明的再一个特点是，鉴定了在鼠耳芥属中编码与许多生物的 1-脱氧-D-木酮糖 5-磷酸还原异构酶 (reductoisomerase) 具有序列相似性的蛋白质的基因，本文中称作 4036 基因，其中所述生物包括集胞蓝细菌属 (*Synechocystis*) (SWISS-PROTQ55663)、枯草芽孢杆菌 (*Bacillus subtilis*) (SWISS-PROT 031753)、和大肠杆菌 (*Escherichia coli*) (SWISS-PROT P45568) (Takahashi 等 (1998) 美国国家科学院院刊, 95: 9879-9884)。本发明的一个重要而出乎意料的特点是，发现该 4036 基因是幼苗生长和发育所必需的。本发明的一个优点是，该新发现的必需基因包含一种新的除草剂作用方式，其使得本领域的技术人员能够容易而快速地鉴定出新的除草剂。

本发明的一个目标是为开发具有除草剂活性的化合物提供植物中的必需基因。遗传结果显示，当鼠耳芥属中 245 基因、5283 基因、2490 基因、3963 基因或 4036 基因发生突变时，在纯合状态所获表型是幼苗致死的。这提示由该突变基因编码的基因产物具有着关键性作用。

采用 T-DNA 插入诱变，本发明的发明者已阐明，由鼠耳芥属 245 基因、鼠耳芥属 5283 基因、鼠耳芥属 2490 基因、鼠耳芥属 3963 基因或 4036

基因编码的活性（本文中称作 245、5283、2490、3963 或 4036 活性）在鼠耳芥属幼苗中是必需的。这意味着，抑制植物中该蛋白质功能的化学制品很可能对植物产生有害影响，并是潜在的良好除草剂候选物。因此本发明提供应用下述基因序列所编码的纯化的蛋白质鉴定其抑制剂的方法，该抑制剂然后可以用作除草剂以抑制例如农作物生长田间不需要植物的生长，所述农作物尤其是农业经济上重要的农作物例如玉米和其它谷类作物例如小麦、黑麦、燕麦、高粱、稻、大麦、黍、草皮草和饲料草等等，以及棉花、甘蔗、甜菜、产油料种子的欧洲油菜和大豆。

本发明公开了来源于鼠耳芥属的命名为 245 基因的核苷酸序列。该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:1，而相应的氨基酸序列显示于 SEQ ID NO:2。其部分基因组 DNA 序列的核苷酸序列显示于 SEQ ID NO:12。本发明还包括与 SEQ ID NO:1 所示核苷酸序列基本相似的核苷酸序列。本发明还包括其氨基酸序列与 SEQ ID NO:2 所示氨基酸序列基本相似的植物蛋白质。该蛋白质能够在鉴定是潜在除草剂的抑制剂的筛选试验中应用。

本发明还公开了来源于鼠耳芥属的命名为 5283 基因的核苷酸序列。该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:3，而相应的氨基酸序列显示于 SEQ ID NO:4。其基因组 DNA 序列的核苷酸序列显示于 SEQ ID NO:14。本发明还包括与 SEQ ID NO:3 所示核苷酸序列基本相似的核苷酸序列。本发明还包括其氨基酸序列与 SEQ ID NO:4 所示氨基酸序列基本相似的植物蛋白质。该蛋白质能够在鉴定是潜在除草剂的抑制剂的筛选试验中应用。

本发明还公开了来源于鼠耳芥属的命名为 2490 基因的核苷酸序列。该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:5，而相应的氨基酸序列显示于 SEQ ID NO:6。其基因组 DNA 序列的核苷酸序列显示于 SEQ ID NO:19。本发明还包括与 SEQ ID NO:5 所示核苷酸序列基本相似的核苷酸序列。本发明还包括其氨基酸序列与 SEQ ID NO:6 所示氨基酸序列基本相似的植物蛋白质。该蛋白质能够在鉴定是潜在除草剂的抑制剂的筛选试验中应用。

本发明还公开了来源于鼠耳芥属的命名为 3963 基因的核苷酸序列。该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:7, 而相应的氨基酸序列显示于 SEQ ID NO:8。其基因组 DNA 序列的核苷酸序列显示于 SEQ ID NO:24, 它含有来自标注为 MDK4.6 的 MDK4 克隆部分和基于本发明人报道 cDNA 克隆的 3'末端添加序列的基因组 DNA 序列。本发明还包括与 SEQ ID NO:7 所示核苷酸序列基本相似的核苷酸序列。本发明还包括其氨基酸序列与 SEQ ID NO:8 所示氨基酸序列基本相似的植物蛋白质。该蛋白质能够在鉴定是潜在除草剂的抑制剂的筛选试验中应用。

本发明还公开了来源于鼠耳芥属的命名为 4036 基因的核苷酸序列。该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:9, 而相应的氨基酸序列显示于 SEQ ID NO:10。其基因组 DNA 序列的核苷酸序列显示于 SEQ ID NO:27。通过来自于栽培品种 Landsberg 的 cDNA 克隆和来自于栽培品种 Columbia 的基因组序列的比较, 观察到 13 个核苷酸差异; 下面的表 1 进一步明确了这些差异。除了这 13 个核苷酸差异外, SEQ ID NO:28 与 SEQ ID NO:9 相同。SEQ ID NO: 28 的相应氨基酸序列显示于 SEQ ID NO: 29。本发明还包括与 SEQ ID NO:9 所示核苷酸序列基本相似的核苷酸序列。本发明还包括其氨基酸序列与 SEQ ID NO:10 和 SEQ ID NO:29 所示氨基酸序列基本相似的植物蛋白质。该蛋白质能够在鉴定是潜在除草剂的抑制剂的筛选试验中应用。

在一个优选的实施方案中, 本发明涉及用于鉴定在植物中能够抑制 245、5283、2490、3963 或 4036 活性的化学制品的方法, 优选包括步骤: a) 获得转基因植物、植物组织、植物种子或植物细胞, 优选稳定转化的转基因植物、植物组织、植物种子或植物细胞, 其含有编码具有 245、5283、2490、3963 或 4036 活性的酶的非天然核苷酸序列并能够超量表达具有酶活性的 245、5283、2490、3963 或 4036 (全长的或截短但仍有活性的) 基因产物; b) 给该转基因植物、植物细胞、组织或部分和等基因非转化植物、植物细胞、组织或部分施用化学制品; c) 测量在施用该化学制品之后该转基因和非转化植物、植物细胞、组织的生长和生活力; d) 比较在施用化学制品后转基因和非转化植物、植物细胞、组织的生长和生活

力；和 e) 选择抑制非转化植物、植物细胞、组织或部分的生活力或生长，但不显著抑制该等基因的转基因植物、植物细胞、组织或部分的生活力或生长的化学制品。在一个优选的实施方案中，具有 245、5283、2490、3963 或 4036 活性的酶由来源于植物，优选鼠耳芥的核苷酸序列编码，期望是该核苷酸序列分别与 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 所示核苷酸序列一致或基本相似。在另一个实施方案中，具有 245、5283、2490、3963 或 4036 活性的酶分别由能够编码 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10 的氨基酸序列的核苷酸序列所编码。在再一个实施方案中，具有 245、5283、2490、3963 或 4036 活性的酶分别具有与 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10 所示氨基酸序列一致或基本相似的氨基酸序列。

本发明还包含具有修饰的 245、5283、2490、3963 或 4036 活性，并因此耐受正常对天然存在的 245、5283、2490、3963 或 4036 活性具有抑制性的除草剂水平之抑制的植物、植物组织、植物种子和植物细胞。本发明所包含的除草剂耐受性植物包括那些本来将会是正常抑制性除草剂的潜在靶标的植物，尤其是以上提及的农业经济上重要的作物。根据该实施方案，用含有与编码修饰的 245、5283、2490、3963 或 4036 基因的核苷酸编码序列可操作地连接、并在植物中起作用的适合启动子的重组 DNA 分子转化，优选稳定转化植物、植物组织、植物种子或植物细胞，其中所述修饰的基因耐受正常情况下抑制野生型未修饰 245、5283、2490、3963 或 4036 基因产物活性的浓度的除草剂的抑制。还可以通过给植物提供多拷贝的野生型 245、5283、2490、3963 或 4036 基因，或通过较野生型强的启动子控制下的野生型 245、5283、2490、3963 或 4036 基因的超量表达，增加除草剂敏感的野生型 245、5283、2490、3963 或 4036 蛋白质的表达，来赋予植物修饰的 245、5283、2490、3963 或 4036 活性。然后通过常规筛选技术对由此产生的转基因植物、植物组织、植物种子或植物细胞进行筛选，藉此分离、分析并开发出除草剂耐受性株系。作为替代方法，可以应用随机或位点特异性诱变产生除草剂耐受性株系。

因此, 本发明提供用含有分离自植物编码具有 245、5283、2490、3963 或 4036 活性之酶的核苷酸序列的 DNA 分子转化的植物、植物细胞、植物种子或植物组织, 其中该 DNA 表达 245、5283、2490、3963 或 4036 活性, 并且其中该 DNA 分子赋予该植物、植物细胞、植物种子、或植物组织对正常抑制天然存在的 245、5283、2490、3963 或 4036 活性的除草剂量的耐受性。根据该实施方案的一个实施例, 具有 245、5283、2490、3963 或 4036 活性的酶分别由与 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 中所示核苷酸序列一致或基本相似的核苷酸序列编码, 或分别具有与 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10 中所示氨基酸序列一致或基本相似的氨基酸序列。

本发明还提供用于抑制植物生长的方法, 包括步骤: 给该植物施用抑制植物中天然存在的 245、5283、2490、3963 或 4036 活性的化学制品。在一个相关方面, 本发明指向用于在含有许多种植作物种子或植物的田间选择性地抑制不需要植物的生长的方法, 包括步骤: (a) 任选地种植除草剂耐受性农作物或农作物种子, 它们是对抑制天然存在的 245、5283、2490、3963 或 4036 活性的除草剂具有耐受性的植物或植物种子; 和 (b) 向田间的这些除草剂耐受性农作物或农作物种子以及不需要植物施用除草剂, 施用量为抑制天然存在的 245、5283、2490、3963 或 4036 活性的除草剂量, 其中该除草剂抑制杂草的生长, 但不显著抑制农作物的生长。

在对本发明的以下描述和非限制性实施例进行研究之后, 本领域的技术人员将明了本发明的其它目标和优点。

正如以下实施例所显示的, 新基因结构的鉴定, 以及 245 基因、5283 基因、2490 基因、3963 基因或 4036 基因对于正常的植物生长和发育的重要性, 采用 T-DNA 插入诱变在鼠耳芥属中首次获得阐明。由于已经确立了 245、5283、2490、3963 或 4036 功能在植物中的重要性, 并已鉴定了编码这些必需活性的基因, 由此发明人提供一种用于新除草剂开发的重要搜寻工具。

鉴定分离幼苗致死突变的鼠耳芥属插入突变株是鉴定必需基因的第一步。用收集自在其基因组中含有 T-DNA 插入片断的单一 T1 植株的 T2 种子作为起始材料，鉴定分离出纯合幼苗致死突变的那些株系。这些株系可以通过以下步骤发现：将种子置于含有杀真菌剂苯菌灵和 maxim 的基本植物生长培养基上，并于室温光照下培养 7 天和 14 天之后鉴别不能存活的幼苗。不能存活的表型包括改变的色素形成或改变的形态。这些表型可以直接在培养皿中或在幼苗移植后在土壤中观察到。

当一个株系被鉴定为分离幼苗致死，确定 T-DNA 中的抗性标记是否与该致死性共分离 (Errampalli 等 (1991) 植物细胞 (The Plant Cell), 3: 149 - 157)。共分离分析通过这些种子置于含有选择剂的培养基上，并给幼苗对于该试剂的抗性或敏感性进行评分来实现。应用的选择剂的例子有潮霉素或磷丝菌素。将大约 35 株抗性幼苗移植至土壤中，并在其后代中检测幼苗致死的分离。在 T-DNA 插入破坏必需基因的情况中，在每一个植株上都会出现该抗性表型与该幼苗致死表型的共分离。因此，在这种情况下，所有抗性植株在下一代中分离幼苗致死；这个结果表明每一个抗性植株对于引起这两个表型的 DNA 而言都是杂合的。

对于表现出 T-DNA 抗性标记与幼苗致死表型共分离的那些株系，进行 Southern 分析作为对每个插入片断的分子性质进行定性的最初步骤。Southern 采用分离自杂合子的基因组 DNA 以及能够与 T-DNA 载体 DNA 杂交的探针来进行。应用该 Southern 分析的结果，选择适当的限制性酶进行质粒拯救，以分子克隆位于 T-DNA 插入片断一侧或两侧的鼠耳芥属基因组 DNA。通过该方法获得的质粒通过限制性酶消化进行分析，以便根据质粒的消化图谱将它们分类。对于每一种类型的质粒克隆，确定其 DNA 序列。分析所获得的序列是否存在非 T-DNA 载体的序列。如果发现这样的序列，则将它们用于搜索 DNA 和蛋白质数据库，该搜索采用 BLAST 和 BLAST2 程序进行 (Altschul 等 (1990) 分子生物学杂志 (J. Mol. Biol.) 215:403-410; Altschul 等 (1997) 核酸研究 25: 3389 - 3402)。通过标准的分子生物学程序鉴定每个基因的其他基因组和 cDNA 序列。

鼠耳芥属 245 基因是通过从标记的幼苗致死系 # 245 分离 T-DNA 边

界侧翼的 DNA 鉴定的。位于 T-DNA 边界侧翼的鼠耳芥属 DNA 区域与基因组调查序列 F17K7TR (索取号 # B24357) 99% 一致。本发明人第一次阐明了该 245 基因产物是植物的正常生长和发育所必需的, 并通过蛋白质同源性明确了 245 基因产物的功能。本发明公开了鼠耳芥属 245 基因的 cDNA 核苷酸序列, 以及鼠耳芥属 245 蛋白质的氨基酸序列。相应于该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:1, 而编码该蛋白质的氨基酸序列显示于 SEQ ID NO:2。在 SEQ ID NO:1 3' 发现的 UTR 显示于 SEQ ID NO:13。相应于其部分基因组 DNA 的核苷酸序列显示于 SEQ ID NO:12。本发明还包括来源于植物的分离氨基酸序列, 其中所述氨基酸序列与 SEQ ID NO:1 中所示核苷酸序列编码的氨基酸序列一致或基本相似, 其中所述氨基酸序列具有 245 活性。采用 BLAST 和 BLAST2 程序以及默认的设置值, 245 基因的序列表现出与多种原核种类的肽释放因子 2 相似。显著的相似种类包括: 大肠杆菌 (RF-2) [Swiss-Prot 索取号 # P07012]; 鼠伤寒沙门氏菌 (*Salmonella typhimurium*) (RF-2 Salty) [Swiss-Prot 索取号 # P28353]; 和结核分枝杆菌 (*Mycobacterium tuberculosis*) (RF-2:prfB) [Swiss-Prot 索取号 # 005782]。对以下蛋白质序列和 245 基因采用 GAP 分析, 获得以下与 245 蛋白质的序列一致性: 大肠杆菌 (RF-2) [Swiss-Prot 索取号 # P07012] (27.2% 一致); 鼠伤寒沙门氏菌 (RF-2 Salty) [Swiss-Prot 索取号 # P28353] (24.6% 一致); 和结核分枝杆菌 (RF-2:prfB) [Swiss-Prot 索取号 # 005782] (27.2% 一致)。此外, 集胞蓝细菌属 (GenPept 索取号 # BAA18577) (31.5% 一致); 和鼠耳芥第 5 号染色体的 P1 克隆 MAB16 (索取号 # AB018112NID) (46.2% 一致)。

鼠耳芥属 5283 基因是通过从标记的幼苗致死系 # 5283 分离 T-DNA 边界侧翼的 DNA 鉴定的。位于 T-DNA 边界侧翼的该鼠耳芥属 DNA 区域与鼠耳芥属一个已测序的 BAC 的内部区域 (BAC T13D8, 1 号染色体) 一致。该 BAC 克隆含有 116,177bp 序列, 其中一个非常小的部分相应于含有该 5283 基因的基因组区域。尽管已有此 BAC 的信息, 但本发明人第一次阐明了该 5283 基因产物是植物的正常生长和发育所必需的, 并通过蛋白质同源性明确了 5283 基因产物的功能。本发明公开了鼠耳芥属 5283 基因

的 cDNA 核苷酸序列，以及鼠耳芥属 5283 蛋白质的氨基酸序列。相应于该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:3，而编码该蛋白质的氨基酸序列显示于 SEQ ID NO:4。相应于其基因组 DNA 的核苷酸序列显示于 SEQ ID NO:14。相应于该 cDNA 序列的 5'UTR 的核苷酸序列显示于 SEQ ID NO:17，而相应于该 cDNA 序列的 3'UTR 的核苷酸序列显示于 SEQ ID NO:18。本发明还包括来源于植物的分离氨基酸序列，其中所述氨基酸序列与 SEQ ID NO:3 中所示核苷酸序列编码的氨基酸序列一致或基本相似，其中所述氨基酸序列具有 5283 活性。采用 BLAST 和 BLAST2 程序以及默认的设置值，5283 蛋白质的序列表现出与如下序列的相似性：粟酒裂殖酵母的 SPBC119.13c [GENPEPT 索取号 # CAA17928]；植物的 SAR DNA 结合蛋白 [SARBP-1: Genbank 索取号 # AF061962 和 SARBP - 2: Genbank 索取号 # AF061963]；和酿酒酵母的 prp31 和 SIK1p (Nop56) [PRP31: Swiss Prot 索取号 # Q12460]。采用 GAP 分析以下蛋白质序列和 5283 蛋白质，获得以下与 5283 蛋白质的序列一致性：粟酒裂殖酵母的 SPBC119.13c [GENPEPT 索取号 # CAA17928] (40.5%一致)；植物的 SAR DNA 结合蛋白 [SARBP-1: Genbank 索取号 # AF061962 (23.5%一致) 和 SARBP - 2: Genbank 索取号 # AF061963 (24.2%一致)]；和酿酒酵母的 prp31 和 SIK1p (Nop56) [PRP31: Swiss Prot 索取号 # Q12460] (24.1%一致)。此外，鼠耳芥 (GenPept 索取号 # AAC18800) 与 5283 蛋白质获得 73.8% 的一致性。

鼠耳芥属 2490 基因是通过从标记的幼苗致死系 # 2490 分离 T-DNA 边界侧翼的 DNA 鉴定的。位于 T-DNA 边界侧翼的该鼠耳芥属 DNA 区域与鼠耳芥属一个已测序的 P1 克隆的内部区域 (P1 MTG13, 第 5 号染色体) 一致。该 P1 克隆含有 50,641bp 序列，其中的一小部分相应于含有该 2490 基因的基因组区域。通过测定 144K24 EST 克隆 (获自 Michigan 州立大学) 的序列，获得含有 2490 蛋白质完整编码序列的 2490 cDNA 序列。尽管已有此 BAC 和 EST 的序列信息，但本发明人第一个明确地确立了该完整基因序列，并阐明了该 2490 基因产物是植物的正常生长和发育所必需的，并且通过蛋白质同源性明确了 2490 基因产物的功能。本发明公开了鼠耳芥属 2490 基因的 cDNA 核苷酸序列，以及鼠耳芥属 2490 蛋白质的氨

氨基酸序列。相应于该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:5, 而编码该蛋白质的氨基酸序列显示于 SEQ ID NO:6。在 SEQ ID NO:5 的 5'端发现的 UTR 序列显示于 SEQ ID NO:20, 而在 SEQ ID NO:5 的 3'端发现的 UTR 序列显示于 SEQ ID NO:21。相应于其基因组 DNA 的核苷酸序列显示于 SEQ ID NO:19。本发明还包括来源于植物的分离氨基酸序列, 其中所述氨基酸序列与 SEQ ID NO:5 中所示核苷酸序列编码的氨基酸序列一致或基本相似, 其中所述氨基酸序列具有 2490 活性。采用 BLAST 和 BLAST2 程序以及默认的设置值, 2490 蛋白质的序列表现出与欧洲油菜的叶绿体被膜 Toc36 (bce42B) (Ko 等 (1995) 生物化学杂志 270: 28601 - 28608; Wu 等 (1994) 生物化学杂志 269: 32264 - 32271; Pang 等 (1997) 生物化学杂志 272: 25623 - 25627) 的相似性。采用 GAP 分析 2490 蛋白和欧洲油菜的叶绿体被膜蛋白 Toc36 (bce42B) (Genbank 索取号 # X79091), 获得与 2490 蛋白质的 81.7% 一致性。

鼠耳芥属 3963 基因是通过从标记的幼苗致死系 # 3963 分离 T-DNA 边界侧翼的 DNA 鉴定的。位于 T-DNA 边界侧翼的该鼠耳芥属 DNA 区域与第 5 号染色体的 P1 克隆 MDK4 (Genbank 索取号 AB010695) 中的基因组序列 100% 一致。本发明人第一个阐明了该 3963 基因产物是植物的正常生长和发育所必需的, 并通过蛋白质同源性明确了 3963 基因产物的功能。本发明公开了鼠耳芥属 3963 基因的 cDNA 核苷酸序列, 以及鼠耳芥属 3963 蛋白质的氨基酸序列。相应于该 cDNA 克隆的核苷酸序列显示于 SEQ ID NO:7, 而编码该蛋白质的氨基酸序列显示于 SEQ ID NO:8。在 SEQ ID NO:7 的 3'端发现的 UTR 序列显示于 SEQ ID NO:26。相应于其基因组 DNA 的核苷酸序列显示于 SEQ ID NO:24。本发明还包括来源于植物的分离氨基酸序列, 其中所述氨基酸序列与 SEQ ID NO:7 中所示核苷酸序列编码的氨基酸序列一致或基本相似, 其中所述氨基酸序列具有 3963 活性。采用 BLAST 和 BLAST2 程序以及默认的设置值, 3963 基因序列表现出与许多 DNA 修复蛋白质的相似性, 这些 DNA 修复蛋白质包括粟酒裂殖酵母的 Rad32p (Genbank 索取号 Q09683); 人的 hMre11 (Genbank 索取号 U37359); 和酿酒酵母的 Mre11p (Genbank 索取号 U60829)。采用 GAP 分析以下蛋白

质序列和 3963 蛋白质, 获得以下与 5283 蛋白质的序列一致性: 粟酒裂殖酵母的 Rad32p (Genbank 索取号 Q09683) (37.5%一致); 人的 hMre11 (Genbank 索取号 U37359) (39.4%一致); 和酿酒酵母的 Mre11p (Genbank 索取号 U60829) (34.7%一致)。

鼠耳芥属 4036 基因是通过从标记的幼苗致死系 # 4036 分离 T-DNA 边界侧翼的 DNA 鉴定的。位于 T-DNA 边界侧翼的该鼠耳芥属 DNA 区域与来自鼠耳芥属第 5 号染色体的 P1 克隆 MQB2 的公开基因组序列 (Genbank 索取号 # AB009053) 100% 一致。本发明人第一个阐明了该 4036 基因产物是植物的正常生长和发育所必需的, 并通过蛋白质同源性明确了 4036 基因的功能。本发明公开了鼠耳芥属 4036 基因的 cDNA 编码核苷酸序列, 以及鼠耳芥属 4036 蛋白质的氨基酸序列。相应于栽培品种 Landsberg 的 cDNA 以及两个栽培品种的 cDNA 的核苷酸序列分别显示于 SEQ ID NO:9 和 SEQ ID NO:28。编码这些蛋白质的相应氨基酸序列显示于 SEQ ID NO:10 和 SEQ ID NO:29。相应于该基因组 DNA 的核苷酸序列显示于 SEQ ID NO:27。通过比较来源于栽培品种 Lansberg 的 cDNA 克隆和来源于栽培品种 Columbia 的基因组序列, 观察到 13 个核苷酸序列差异, 这些变异列于下表 1 中。

表 1. 在来源于栽培品种 Lansberg 的 4036 cDNA 克隆和来源于栽培品种 Columbia 的 4036 基因组序列之间所观察到的核苷酸差异

核苷酸 # *	栽培品种 Landsberg	栽培品种 Columbia	含有核苷酸差异的密码子 (栽培品种 Landsberg 中的氨基酸残基和栽培品种 Columbia 中的氨基酸残基)**
115	G	A	GAT 到 AAT (Asp 到 Asn)
207	T	C	GTT 到 GTC (Val 到 Val)
273	C	T	TCC 到 TCT (Ser 到 Ser)
276	C	T	ATC 到 ATT (Ile 到 Ile)
321	T	C	TTT 到 TTC (Phe 到 Phe)
393	G	A	GCG 到 GCA (Ala 到 Ala)
485	T	A	CTA 到 CAA (Leu 到 Gln)
464	C	T	CCC 到 CTC (Pro 到 Leu)
559	A	C	AAG 到 CAG (Lys 到 Gln)

963	T	G	CCT 到 CCG (Pro 到 Pro)
1101	T	A	CCT 到 CCA (Pro 到 Pro)
1254	T	C	TTT 到 TTC (Phe 到 Phe)
1393	G	A	GAT 到 AAT (Asp 到 Asn)

*SEQ ID NO:9 用作核苷酸编号的参考

**氨基酸残基: Ala(丙氨酸); Asn(天冬酰胺); Asp(天冬氨酸); Gln(谷氨酰胺); Ile(异亮氨酸); Leu(亮氨酸); Lys(赖氨酸); Phe(苯丙氨酸); Pro(脯氨酸); Ser(丝氨酸); 和 Val(缬氨酸)

本发明还包括来源于植物的分离氨基酸序列, 其中所述氨基酸序列与 SEQ ID NO:9 中所示核苷酸序列所编码的氨基酸序列一致或基本相似, 其中所述氨基酸序列具有 4036 活性。采用 BLAST 和 BLAST2 程序以及默认的设定值, 4036 基因序列表现出与多种生物的 1-脱氧-D-木酮糖 5-磷酸还原异构酶的相似性, 这些生物包括集胞蓝细菌属 (SWISS-PROTQ55663)、枯草芽孢杆菌 (SWISS - PROT 031753)、和大肠杆菌 (SWISS-PROT P45568) (Takahashi 等 (1998) 美国国家科学院院刊 95: 9879 - 9884)。采用 GAP 分析以下蛋白质序列和 4036 蛋白质, 获得以下与 4036 蛋白质的序列一致性: 集胞蓝细菌属的 1-脱氧-D-木酮糖 5-磷酸还原异构酶 (SWISS-PROTQ55663) (66.1%一致)、枯草芽孢杆菌的 1-脱氧-D-木酮糖 5-磷酸还原异构酶 (SWISS - PROT 031753) (45.4%一致)、和大肠杆菌的 1-脱氧-D-木酮糖 5-磷酸还原异构酶 (SWISS-PROT P45568) (Takahashi 等 (1998) 美国国家科学院院刊 95: 9879 - 9884) (44.6%一致)。

为了在宿主生物中重组产生 245、5283、2490、3963 或 4036 活性, 将编码具有 245、5283、2490、3963 或 4036 活性之蛋白质的核苷酸序列插入设计用于所选宿主的表达盒中, 并将其导入待重组产生该活性的宿主细胞中。例如, SEQ ID NO:1 或与作为 3'UTR 的 SEQ ID NO:13 相连的 SEQ ID NO:1、与 SEQ ID NO:1 基本相似的核苷酸序列、或 245 编码序列的同系物可以用于重组产生具有 245 活性的蛋白质。对于适合所述宿主的特定调节序列例如启动子、信号序列、5'和 3'非翻译序列、和增强子的选择属于本领域的常规技术水平范围内。可以将含有按正确阅读框可

操作地连接在一起的单个元件的所获分子插入能够被转化至宿主细胞的载体中。对于宿主生物例如大肠杆菌、酵母和昆虫细胞，以及杆状病毒表达载体例如来源于苜蓿银纹夜蛾 (*Autographica californica*) 核型多角体病毒 (AcMNPV) 基因组的载体，用于重组产生蛋白质的适合表达载体和方法是熟知的 (见例如 Luckow 和 Summers, *Bio/Technol.* 6:47 (1988)). 优选的杆状病毒/昆虫系统是 pAcHLT (Pharming, San Diego, CA), 用于存在线性苜蓿银纹夜蛾杆状病毒 DNA (Pharmigen, San Diego, CA) 时转染草地夜蛾 (*Spodoptera frugiperda*) sf9 细胞 (ATCC)。所获病毒用于转染 HighFive *Tricoplusia ni* 细胞 (Invitrogen, La Jolla, CA)。以相似方式获得 5283、2490、3963 或 4036 活性的重组产生。

在一个优选实施方案中，编码具有 245、5283、2490、3963 或 4036 活性之蛋白质的核苷酸序列来源于真核生物，例如哺乳动物、蝇或酵母，但优选来源于植物。在再一个优选实施方案中，该核苷酸序列分别与 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 中所示核苷酸序列一致或基本相似，或分别编码具有 245、5283、2490、3963 或 4036 活性的蛋白质，其氨基酸序列分别与 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10 中所示氨基酸序列一致或基本相似。SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 中所示核苷酸序列编码鼠耳芥属 245 蛋白质、鼠耳芥属 5283 蛋白质、鼠耳芥属 2490 蛋白质、鼠耳芥属 3963 蛋白质或鼠耳芥属 4036 蛋白质，其氨基酸序列分别显示于 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10。在另一个优选实施方案中，该核苷酸序列来源于原核生物，优选细菌例如大肠杆菌。

可以采用各种标准技术分离和纯化重组制备的具有 245、5283、2490、3963 或 4036 活性的蛋白质。可以采用的实际技术将取决于所选宿主生物、蛋白质是否设计为分泌的、以及本发明技术人员所熟知的其它因素 (见例如 Ausubel, F. 等, “当代分子生物学实验指南 (Current Protocols in Molecular Biology)”, 第 16 章, John Wiley & Sons 公司出版 (1994))。

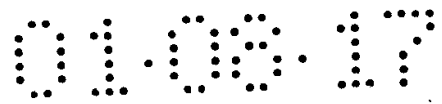
重组产生的具有 245、5283、2490、3963 或 4036 活性的蛋白质可以

用于多种目的。例如，它们可以用于体外分析以对未鉴定出其靶标的已知除草剂化学制品进行筛选，以确定其是否对 245、5283、2490、3963 或 4036 活性产生抑制。这样的体外分析还可以用作更一般的筛选，以鉴定抑制该酶活性并由此成为新的除草剂候选分子的化学制品。或者，可以应用重组产生的具有 245、5283、2490、3963 或 4036 活性的蛋白质，阐明这些分子的复杂结构并进一步分析它们与抑制剂结合的特点，以便合理地设计新的抑制性除草剂以及酶的除草剂耐受形式。

体外抑制剂分析：发现分别与 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 的基因产物相互作用的小分子配体

一旦一个蛋白质已被鉴定为潜在的除草剂靶标之后，下一步是开发允许对大量化学制品进行筛选的分析试验，以确定与该蛋白质相互作用的化学制品。尽管开发针对已知功能的蛋白质的分析试验是简单的，但开发关于未知功能的蛋白质的分析试验则较为困难。该难题可通过使用能在不了解蛋白质的生物学功能时检测蛋白质与化合物之间相互作用的技术来克服。本文给出了三种方法的简短描述，它们包括荧光相关光谱学、表面增强激光解吸/电离 (surface-enhanced laser desorption/ionization)、和 biacore 技术。

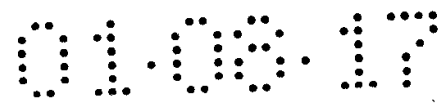
荧光相关光谱学 (FCS) 的理论出现于 1972 年，但仅在近年才有了实施 FCS 的技术 (Madge 等, (1972) Phys. Rev. Lett., 29: 705-708; Malti 等 (1997) 美国国家科学院院刊, 94: 11753 - 11757)。FCS 测量在小样品体积中荧光分子的平均扩散速率。该样品的大小可以低至 10^3 个荧光分子，该样品的体积可以低至单个细菌的细胞质。扩散速率是分子质量的函数，并随着质量的增加而降低。因此可以将 FCS 应用于蛋白质-配体的相互作用分析，方法是测量结合导致的质量改变以及由此引起的分子扩散速率的改变。在一个典型实验中，将待分析的目标表达成在 N 或 C 端插有顺序标记物例如聚组氨酸序列的重组蛋白质。该表达在大肠杆菌、酵母或昆虫细胞中进行。通过层析纯化该蛋白质。例如，可以利用聚组氨酸标记物使表达的蛋白质与金属螯合柱例如螯合在亚氨基二乙酸琼脂糖上的 Ni^{2+} 结合。然后用荧光标记物例如羧基四甲基罗丹明或 BODIPY®



(Molecular Probes, Eugene, OR) 标记该蛋白质。之后将该蛋白质在溶液中暴露给可能的配体, 并采用可获自 Carl Zeiss 公司 (Thornwood, NY) 的仪器通过 FCS 测量其扩散速率。通过该蛋白质扩散速率的改变确定配体的结合。

表面增强激光解吸/电离 (SELDI) 是 Hutchens 和 Yip 在 20 世纪 80 年代后期发明的 (Hutchens 和 Yip (1993) *Rapid Commun. Mass Spectrom.* 7: 576-580)。当与飞行时间质谱仪 (TOF) 偶联时, SELDI 提供了一种快速分析保留在芯片上的分子的方法。通过将靶蛋白质共价结合在芯片上, 并通过 MS 分析与该蛋白质结合的小分子, 能够将 SELDI 应用于配体-蛋白质相互作用分析 (Worrall 等 (1998) *分析生物化学 (Anal. Biochem.)* 70: 750-756)。在一个典型的实验中, 按用于 FCS 的所述方式表达待分析的目标。然后将纯化的蛋白质用于该分析, 而不需要进一步制备。通过利用聚组氨酸标记物或通过其它相互作用例如离子交换或疏水相互作用, 将该蛋白质与 SELDI 芯片结合。然后通过例如能够以顺序方式吸取配体的递送系统 (自动进样器), 将由此制备的芯片暴露给可能的配体。然后对该芯片进行递增严紧度的洗涤, 例如用含有递增离子强度的缓冲溶液进行一系列洗涤。每次洗涤之后, 通过将该芯片送至 SELDI-TOF, 分析结合的物质。特异地与靶标结合的配体将通过需要洗脱它们的洗涤严紧性来鉴定。

Biacore 以固定在表层的蛋白质与配体的结合导致的表层反射率的改变为基础。在该系统中, 将大量小配体顺序地注入带有固定化蛋白质的一个 $2-5 \mu\text{l}$ 的小室中。通过表面等离子共振 (surface plasmon resonance, SPR) 记录该表面反射的激光, 以检测结合。一般, 针对表面层的给定质量浓度变化的反射率改变实际上对于所有蛋白质和肽都是相同的, 这就允许可以将一个单一的方法应用于任何蛋白质 (Liedberg 等 (1983) *Sensors Actuators* 4:299-304; Malmquist (1993) *自然 (Nature)* 361:186-187)。在一个典型的实验中, 按用于 FCS 的所述方式表达待分析的目标。然后无需进一步制备即可将该纯化蛋白质用于该分析。通过利用聚组氨酸标记物或通过其它相互作用例如离子交换或疏水相互作用



用,使该蛋白质与 Biacore 芯片结合。然后通过整合在 Biacore (Uppsala, 瑞典) 出售的仪器中用于以顺序方式吸取配体的递送系统(自动进样器),将由此制备的芯片暴露给可能的配体。记录芯片上的 SPR 信号,反射率的改变反映了固定的靶标与配体之间的相互作用。信号动力学开关速率的分析使得可以区分非特异和特异相互作用。

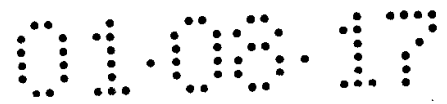
而且,对于与多肽相互作用的小分子配体的分析也是抑制剂分析。例如,用于鉴定必需植物基因例如 245、5283、2490、3963 或 4036 基因的抑制剂的此类抑制剂分析包括步骤:

- a) 在存在如下蛋白质功能的可疑抑制剂的情况下,使植物的 245、5283、2490、3963 或 4036 蛋白质与其底物反应;
- b) 比较存在此可疑抑制剂时的酶活性速度和相同条件下不存在该可疑抑制剂时的酶活性速度; 和
- c) 确定该可疑抑制剂是否抑制了 245、5283、2490、3963 或 4036 蛋白质。

例如,可以通过该分析中 245、5283、2490、3963 或 4036 活性的降低或完全抑制,确定对植物 245、5283、2490、3963 或 4036 蛋白质的抑制性作用。该确定可以通过比较在有和无该候选抑制剂的情况下反应过程中用掉的底物或产生的中间物或产物的量来实现。

在一个实施方案中,将例如通过体外筛选鉴定的可疑除草剂,以不同浓度施用给植物。优选将该可疑除草剂喷洒在植物上。在施用了该可疑除草剂之后,记录其对植物产生的影响,例如死亡或生长抑制。

在另一个实施方案中,针对 245、5283、2490、3963 或 4036 活性之抑制剂的体内筛选方法,应用能够超量表达具有 245、5283、2490、3963 或 4036 活性的核苷酸序列的转基因植物、植物组织、植物种子或植物细胞,其中该 245、5283、2490、3963 或 4036 基因产物在该转基因植物、植物组织、植物种子或植物细胞中具有酶活性。该核苷酸序列优选来自真核生物,例如酵母,但优选来自植物。在一个更优选的实施方案中,该核苷酸序列与 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 中所示核苷酸序列一致或基本相似,或编码具有 245、

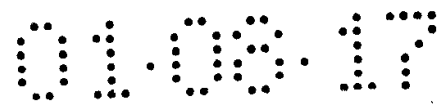


5283、2490、3963 或 4036 活性的酶，其氨基酸序列分别与 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10 中所示氨基酸序列一致或基本相似。在另一个优选实施方案中，该核苷酸序列来源于原核生物，优选细菌例如大肠杆菌。

然后将化学制品施用于该转基因植物、植物组织、植物种子或植物细胞以及等基因的非转基因植物、植物组织、植物种子或植物细胞，并在施用了该化学制品后确定和比较该转基因和非转基因植物、植物组织、植物种子或植物细胞的生长或生活力。选择能够抑制非转基因植物的生长，但又不影响转基因植物的生长的化合物，作为 245、5283、2490、3963 或 4036 活性的抑制剂。

本发明还指向耐受如下除草剂的植物、植物组织、植物种子和植物细胞，该除草剂抑制这些植物中天然存在的 245、5283、2490、3963 或 4036 活性，其中该耐受性分别由改变的 245、5283、2490、3963 或 4036 活性赋予。改变的 245、5283、2490、3963 或 4036 活性可以根据本发明通过增加除草剂敏感的野生型 245、5283、2490、3963 或 4036 基因的表达赋予植物，例如通过提供额外的野生型 245、5283、2490、3963 或 4036 基因，和/或通过例如用强启动子驱动表达的方式分别超量表达内源性 245、5283、2490、3963 或 4036 基因来进行。还可以通过分别在植物中表达与 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 基本相似的核苷酸序列或它们的同系物来获得改变的 245、5283、2490、3963 或 4036 活性。更进一步地，可以通过在植物中分别表达修饰的除草剂耐受性 245、5283、2490、3963 或 4036 基因，赋予该植物改变的 245、5283、2490、3963 或 4036 活性。还可以应用这些技术的组合。代表性植物包括为了达到其通常预期的目的而应用这些除草剂的任何植物。优选农业经济上重要的农作物例如棉花、大豆、产油料种子的欧洲油菜、甜菜、玉米、稻、小米、大麦、燕麦、黑麦、高粱、小米、草皮、饲料草、草皮草等等。

通过增加表达分别实现改变的 245 活性或 5283、2490、3963 或 4036 活性，结果分别导致植物细胞中 245 活性或 5283、2490、3963 或 4036



活性的水平可以至少足以克服由如下施用量的除草剂造成的生长抑制，所述施用量为足以抑制对照植物的正常生长的量。表达的酶水平一般是天然表达量的至少 2 倍，优选至少 5 倍，更优选至少 10 倍。增加的表达可以是由于植物细胞中分别多拷贝的野生型 245 基因或 5283、2490、3963 或 4036 基因；在该基因中编码序列的多次出现(即基因扩增)，或内源性基因的非编码调节序列中的突变造成的。具有该改变的基因活性的植物可以通过本领域已知的方法在植物中直接筛选获得(见例如美国专利 5,162,602 和美国专利 4,761,373，以及其中引用的参考文献)。这些植物还可以通过本领域已知的遗传工程技术获得。除草剂敏感性 245 基因或 5283、2490、3963 或 4036 基因表达的分别增加还可以通过用如下重组或嵌合 DNA 分子转化植物细胞来实现，所述 DNA 分子含有能够在植物细胞中驱动相连的结构基因表达的启动子，而且该启动子可操作地与分别编码 245 蛋白质或 5283、2490、3963 或 4036 蛋白质的同源或异源结构基因或它们的同系物相连。优选地，该转化是稳定的，藉此提供可遗传的转基因性状。

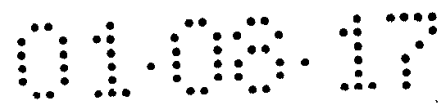
根据该实施方案，用如下重组 DNA 分子稳定转化植物、植物组织、植物种子或植物细胞，该 DNA 分子含有可操作地与分别编码除草剂耐受形式 245、5283、2490、3963 或 4036 蛋白质的编码序列连接的并能在植物中起作用的适合启动子。该酶的除草剂耐受形式具有至少一个氨基酸替代、添加或缺失，它们赋予了对抑制该酶未修饰天然存在形式的除草剂的耐受性。然后通过常规筛选技术筛选由此产生的转基因植物、植物组织、植物种子或植物细胞，藉此分离除草剂耐受株系，并对其进行表征和开发。以下描述了分别用于获得编码除草剂耐受形式 245、5283、2490、3963 或 4036 蛋白质的基因的方法：

一个一般的策略涉及对微生物的直接或间接的诱变过程。例如，可以使用诱变剂例如 UV 光，或者乙基或甲基甲磺酸，对可遗传操作的微生物例如大肠杆菌或酿酒酵母进行体内随机诱变。诱变程序描述于例如 Miller, 分子遗传学实验(Experiments in Molecular Genetics), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1972); Davis

等, 高级细菌遗传学 (Advanced Bacterial Genetics), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1980); Sherman 等, 酵母遗传学方法 (Methods in Yeast Genetics), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1983); 和美国专利 4,975,374. 选择用于诱变的微生物分别含有正常的抑制剂敏感性 245、5283、2490、3963 或 4036 基因, 并依赖于该基因所赋予的活性。在存在对未修饰基因产生抑制的抑制剂浓度时, 生长该诱变细胞。选择存在抑制剂时比未诱变的微生物生长更好的诱变微生物菌落 (即表现出对该抑制剂的抗性), 用于进一步分析。通过克隆或通过 PCR 扩增, 从这些菌落分离出分别赋予对该抑制剂的耐受性的 245、5283、2490、3963 或 4036 基因, 并阐明它们的序列。然后将编码改变的基因产物的序列克隆回该微生物, 以验证它们赋予抑制剂耐受性的能力。

获得植物 245、5283、2490、3963 或 4036 基因的突变除草剂耐受性等位基因的方法, 涉及在植物中进行直接筛选。例如, 可以通过将经本领域已知方法消毒的种子置于含有递增浓度抑制剂的简单极限盐培养基平板上, 确定诱变的 245、5283、2490、3963 或 4036 基因对鼠耳芥属、大豆或玉米等植物生长抑制的影响。所述浓度的范围是百万分之 0.001、0.003、0.01、0.03、0.1、0.3、1、3、10、30、110、300、1000 和 3000 (ppm)。能够重复检测到显著生长抑制的最低剂量被用于随后的实验中。该最低剂量的确定在本领域中是常规的。

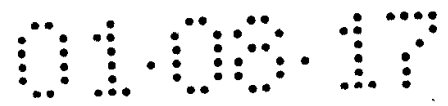
利用植物材料的诱变增加在所选择的群体中抗性等位基因发生的频率。诱变的种子材料来自各种来源, 包括化学或物理诱变的种子, 或化学或物理诱变的花粉 (Neuffer, 用于生物研究的玉米 (Maize for Biological Research) Sheridan 编, Univ. Press, Grand Forks, ND., 第 61-64 页 (1982)), 然后用此花粉对植物受精, 并收集获得的 M_1 突变种子。典型地, 对于鼠耳芥属, 从经化学制品例如乙基甲磺酸或经物理因素例如 γ 射线或快中子诱变的种子生长出的植物获得后代种子, 即 M_2 种子 (Lehie Seeds, Tucson, AZ), 将该种子以高达 10,000 个种子/平板 (直径 10cm) 的密度接种在含有合适抑制剂浓度的极限盐培养基上, 以



对耐受性进行筛选。将接种后持续生长并保持绿色 7-21 天的幼苗移植到土壤中，并使其生长至成熟并结籽。测试这些种子的后代对于该除草剂的耐受性。如果该耐受性性状是显性的，其种子以 3:1/抗性:敏感性的比例分离的植物被认为在 M_2 代是抗性杂合的。产生所有抗性种子的植物被认为在 M_2 代是抗性纯合的。这种在完整种子上进行的诱变并对它们的 M_2 代种子进行筛选还可以在其它物种例如大豆上进行（见例如美国专利 5,084,082）。或者，可以通过用经化学或物理方法诱变的花粉进行受精，获得用于除草剂耐受性筛选的突变种子。

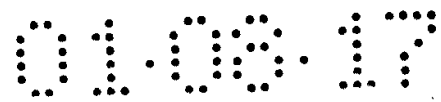
按以下举例说明的方法，验证该除草剂耐受性的遗传基础分别是 245、5283、2490、3963 或 4036 基因。首先，采用 PCR，以及分别基于 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 中所示鼠耳芥属 cDNA 编码序列的引物，更优选基于用来产生耐受性等位基因的植物的未改变 245、5283、2490、3963 或 4036 基因序列的引物，从表现出对该除草剂具有抗性的植物分别分离出该 245、5283、2490、3963 或 4036 基因的等位基因。在对这些等位基因测序以确定编码序列中是否存在突变之后，用这些推测的赋予耐受性的等位基因转化植物，并测试这些等位基因赋予该植物耐受此抑制剂的能力。这些植物可以是鼠耳芥属植物，或可以是其生长分别对 245、5283、2490、3963 或 4036 的抑制剂敏感的任何其它植物。其次，相对于已知的限制性片断长度多态性 (RFLP)（见例如 Chang 等，美国国家科学院院刊 85: 6856-6860 (1988)；Nam 等，植物细胞 (Plant Cell) 1: 699-705 (1989)）、断裂的扩增多态序列 (CAPS) (Konieczny 和 Ausubel (1993) 植物杂志 (The Plant Journal), 4(2): 403-410)、或 SSLPs (Bell 和 Ecker (1994) 基因组 (Genomics), 19: 137-144)，对插入的 245、5283、2490、3963 或 4036 基因进行作图。采用相同的标记独立地对该 245、5283、2490、3963 或 4036 抑制剂耐受性性状分别作图。若耐受性分别是由于 245、5283、2490、3963 或 4036 基因中的突变引起的，则该耐受性性状在图谱上的位置将难以与 245、5283、2490、3963 或 4036 基因的位置区分开来。

获得 245、5283、2490、3963 或 4036 基因的除草剂耐受性等位基因



的另一种方法是在植物细胞培养物中进行筛选。将目的植物的植物组织例如胚、叶盘等的外植体，或活跃生长的愈伤组织或悬浮培养物，生长在含有递增浓度的抑制性除草剂或适用于实验室环境的类似抑制剂的培养基上。记录不同培养物的生长改变程度。在某些培养物中，出现出甚至在有正常抑制性浓度的抑制剂情况下也能持续生长的快速生长变异集落。这种较快的生长变异体的发生频率能够通过将该组织或细胞暴露给抑制剂之前用化学或物理诱变剂进行处理来提高。按前面段落中描述的方法分别分离并检测该 245、5283、2490、3963 或 4036 基因的推测赋予耐受性的等位基因。然后可以改造那些被鉴定为赋予除草剂耐受性的等位基因，以获得最佳表达，并用其转化植物。或者，可以从含有这些等位基因的组织或细胞培养物再生植物。

再一个方法涉及在细菌或酵母中分别对野生型除草剂敏感植物的 245、5283、2490、3963 或 4036 基因进行诱变，之后在含有抑制性浓度（即足以造成异常生长、抑制生长或引起细胞死亡的浓度）的该抑制剂的培养基上培养该微生物，然后筛选在该抑制剂存在时正常生长的那些集落。更具体地，将植物 cDNA，例如分别编码 245、5283、2490、3963 或 4036 蛋白质的鼠耳芥属 cDNA，克隆至本来缺少该 245、5283、2490、3963 或 4036 活性的微生物中。然后通过本领域已知的几种化学或酶学方法之任意一种，例如亚硫酸氢钠（Shortle 等，酶学方法 (Methods Enzymol.) 100:457-468 (1983)）；甲氧胺 (Kadonaga 等，核酸研究 13: 1733-1745 (1985))；寡核苷酸指导的饱和诱变 (Hutchinson 等，美国国家科学院院刊 83: 710-714 (1986))；或各种聚合酶错误掺入策略 (见例如 Shortle 等，美国国家科学院院刊 79:1588-1892 (1982)；Shiraishi 等，基因 (Gene) 64: 313-319 (1988)；和 Leung 等，技术 (Technique) 1:11-15 (1989))，对该转化的微生物进行体内诱变或体外诱变。挑取在存在正常抑制性浓度的抑制剂时生长正常的菌落，并通过重复再次划线进行纯化。纯化它们的质粒，并通过将其重新转化至分别缺少 245、5283、2490、3963 或 4036 活性的微生物中，测试其赋予对该抑制剂的耐受性的能力。然后确定通过该测试的质粒的 cDNA 插入片段的 DNA 序列。

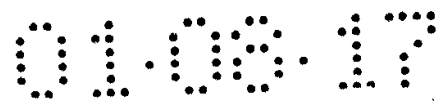


也可以采用涉及体外重组，又称 DNA 改组的方法，分别获得除草剂抗性 245、5283、2490、3963 或 4036 蛋白质。通过 DNA 改组，分别向编码 245、5283、2490、3963 或 4036 活性的核苷酸序列中引入突变，优选随机突变。DNA 改组还导致分别在 245、5283、2490、3963 或 4036 基因内的序列重组和重排，或分别在两个或多个不同的 245、5283、2490、3963 或 4036 基因之间的序列重组或交换。这些方法使得可以分别产生无数的突变 245、5283、2490、3963 或 4036 编码序列。对这些突变基因或改组基因进行筛选，以鉴别期望的性质例如改良的除草剂耐受性，或鉴别提供对不同类型抑制剂化学的广谱耐受性的突变。这些筛选完全属于本领域的常规技术范围之内。

在一个优选的实施方案中，分别从至少一个 245、5283、2490、3963 或 4036 基因模板分别形成诱变的 245、5283、2490、3963 或 4036 基因，其中分别将该模板 245、5283、2490、3963 或 4036 基因断裂成具有期望大小的双链随机片断，并包含步骤：向所获得的该双链随机片断群体加入一或多个单链或双链寡核苷酸，其中所述寡核苷酸含有与该双链随机片断一致的区域和异源的区域；将获得的该双链随机片断和寡核苷酸混合物变性为单链片断；在导致所述单链片断在所述的一致区域退火以形成退火片断对的条件下，将获得的该单链片断群体与聚合酶一起孵育，所述一致区域足以使一对中的一个成员引发另一个的复制，由此形成诱变的双链多核苷酸；再重复第二步和第三步至少两个循环，其中下一个循环第二步中的所获混合物包括来自前一个循环第三步的诱变双链多核苷酸，并且此下一个循环形成进一步诱变的双链多核苷酸，其中该诱变多核苷酸分别是对于分别抑制天然存在的 245、5283、2490、3963 或 4036 活性的除草剂具有增强耐受性的突变 245、5283、2490、3963 或 4036 基因。在一个优选实施方案中，该双链随机片断群体中单个种类双链随机片断的浓度不少于总 DNA 重量的 1%。在又一优选实施方案中，该模板双链多核苷酸含有至少大约 100 种多核苷酸。在另一个优选实施方案中，该双链随机片断的大小从大约 5bp 至大约 5kb。在又一优选实施方案中，该方法的第四个步骤包括重复该第二和第三个步骤至少 10 个循环。该方

法描述于例如 Stemmer 等 (1994) 自然 370: 389-391, 美国专利 5,605,793, 美国专利 5,811,238, 和 Cramer 等 (1998) 自然 391: 288-291, 以及 WO 97/20078, 这些文献以参考文献方式并入本文。

在另一个优选实施方案中, 按例如 Zhao 等 (1998) 自然生物技术 (Nature Biotechnology) 16:258-261 中描述的方法, 通过交错延伸方法 (StEP), 对两个或多个不同的 245 基因的任意组合进行体外诱变。此两或多个 245 基因用作 PCR 扩增的模板, PCR 反应的延伸循环优选在比聚合酶的最佳聚合温度低的温度下进行。以相似的方式, 对 5283、2490、3963 或 4036 基因进行 StEP。例如, 当采用具有大约 72°C 最佳温度的热稳定聚合酶时, 该延伸反应的温度期望低于 72°C, 更期望低于 65°C, 优选低于 60°C, 更优选该延伸反应的温度为 55°C。此外, 该 PCR 循环的延伸反应持续时间期望比本领域通常所用的短, 更期望是少于 30 秒, 优选少于 15 秒, 更优选该延伸反应的持续时间为 5 秒。在每一个延伸反应中仅聚合一个短的 DNA 片断, 使得每一个循环的变性和退火之后可以在起始 DNA 分子之间交换作为延伸产物的模板, 藉此在延伸产物中产生多样性。PCR 反应中循环的最佳数目分别取决于待诱变的 245、5283、2490、3963 或 4036 基因的长度, 但期望大于 40 个循环, 更期望大于 60 个循环, 优选采用 80 个以上的循环。对于 245、5283、2490、3963 或 4036 基因各自的每种组合, 最佳延伸条件和最佳的 PCR 循环数可以按本领域熟知方法进行确定。该 PCR 反应的其它参数与本领域通常采用的基本相同。用于该扩增反应的引物优选设计为能与位于该 245、5283、2490、3963 或 4036 基因之外的 DNA 序列, 例如与分别含有该 245、5283、2490、3963 或 4036 基因的载体的 DNA 序列退火, 藉此分别用于该 PCR 反应的不同 245、5283、2490、3963 或 4036 基因优选包含在不同的载体中。这些引物期望与分别位于 245、5283、2490、3963 或 4036 序列之外不足 500bp, 优选不足 200bp, 更优选分别距离 245、5283、2490、3963 或 4036 序列不足 120bp 的序列退火。优选该 245、5283、2490、3963 或 4036 序列分别被限制性位点围绕, 这些限制性位点包括在 PCR 反应期间扩增的 DNA 序列中, 藉此利于该扩增产物克隆至适合的载体中。在另一个优选的实施方案中, 按 WO



98/05765 中所描述的方法，分别制备带有粘性末端的 245、5283、2490、3963 或 4036 基因片断。可以通过将分别相应于 245、5283、2490、3963 或 4036 基因一部分的第一寡核苷酸与第二寡核苷酸相连，来制备该粘性末端，其中该第二寡核苷酸是该基因中所不存在的，或是与第一寡核苷酸所相应的基因部分不相邻的该基因一部分，并且该第二寡核苷酸含有至少一个核糖核苷酸。采用第一寡核苷酸作为模板，第二寡核苷酸作为引物，制备双链 DNA。切割该核糖核苷酸，并将其除去。还除去位于该核糖核苷酸 5'端的核苷酸，结果获得具有粘性末端的双链片断。这些片断通过连接被随机地重新组装起来，获得新的基因序列组合。

在本发明中，采用任何单个的 245、5283、2490、3963 或 4036 基因，或 245、5283、2490、3963 或 4036 基因的任何组合用于体外重组，例如，分别获自植物例如鼠耳芥的 245、5283、2490、3968 或 4036 基因，例如 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 中分别显示的 245、5283、2490、3963 或 4036 基因，或分别来自大肠杆菌的 245 样、5283 样、2490 样、3963 样或 4036 样基因 (Craigen 等 (1985) 美国国家科学院院刊 82: 3616-3620; Craigen 和 Caskey (1987) Biochimie, 69:1031-1041; Ito 等 (1998) 美国国家科学院院刊 95: 8165-8169; 所有这些文献均以参考文献方式并入本文)。在本发明中采用了单个的完整 245、5283、2490、3963 或 4036 基因或其部分。将分别通过以上描述的方法获得的突变 245、5283、2490、3963 或 4036 基因文库克隆至适合的表达载体中，并用该所获得的载体转化适当宿主，例如衣藻 (chlamydomonas) 等藻类、酵母或细菌。适当的宿主优选本身缺少 245、5283、2490、3963 或 4036 活性的宿主，例如大肠杆菌。将分别用含有 245、5283、2490、3963 或 4036 基因文库的载体转化的宿主细胞培养在含有抑制性浓度之抑制剂的培养基上，并筛选在存在抑制剂时生长的那些菌落。挑取存在正常抑制性浓度的抑制剂时生长的菌落，并通过重复再划线进行纯化。纯化它们的质粒，然后测定通过该测试的质粒中 cDNA 插入片断的 DNA 序列。

可以采用与用于分别鉴定 245、5283、2490、3963 或 4036 活性之抑

制剂的方法（抑制剂分析实验，见上）相同的方法，分别对耐受抑制剂的修饰 245、5283、2490、3963 或 4036 基因进行鉴定，改动如下：首先，在一个反应混合物中分别用突变的 245、5283、2490、3963 或 4036 蛋白质替代抑制剂分析实验中的野生型 245、5283、2490、3963 或 4036 蛋白质。其次，两个反应混合物中都存在野生型酶的抑制剂。第三点，比较突变活性（存在抑制剂和突变酶时的活性）和未突变活性（存在抑制剂和野生型酶时的活性），以确定当与未突变活性比较时，在突变活性中是否观察到酶活性的显著增加。突变活性是存在适合的底物和该抑制剂时以任何方式测量的突变酶活性。未突变活性是存在适合底物和该抑制剂时以任何方式测量的野生型酶活性。

除了用于产生除草剂耐受性植物外，分别编码除草剂耐受性 245、5283、2490、3963 或 4036 蛋白质的基因还可以用作植物细胞转化方法中的筛选标记。例如，经异源 DNA 序列转化的植物、植物组织、植物种子或植物细胞也可以用分别编码改变的 245、5283、2490、3963 或 4036 活性并能够被植物表达的序列转化。将该转化细胞转移至含有该酶抑制剂的培养基上，该抑制剂的量将足以抑制不表达该修饰编码序列的植物细胞的生长或存活，其中仅有转化细胞可以生长。该方法可以应用于能够被修饰的 245、5283、2490、3963 或 4036 基因转化的任何植物细胞，而且可以和任何目的异源 DNA 序列一起使用。该异源 DNA 序列和该修饰基因的表达可以由同一个在植物细胞中起作用的启动子驱动，或由不同的启动子驱动。

可以采用常规重组 DNA 技术分别将 245、5283、2490、3963 或 4036 基因的野生型或除草剂耐受性形式，或它们的同系物掺入植物或细菌细胞中。一般地，这涉及到采用本领域已知的标准克隆程序分别将编码该 245、5283、2490、3963 或 4036 基因的 DNA 分子插入表达系统中，对于该表达系统该 DNA 分子是异源的（即正常并不存在的）。该载体含有使该插入蛋白质编码序列在含有该载体的宿主细胞中转录和翻译的必需元件。本领域已知的大量载体系统都可以使用，例如质粒、噬菌体病毒和其它修饰病毒。还可以对该表达系统的成分进行修饰以增加表达。例如，



可以使用截短序列、核苷酸替代、核苷酸优化或其它的修饰。在适合条件下本领域已知的表达系统均可用于转化实际上任何农作物植物细胞。分别含有野生型或除草剂耐受型 245、5283、2490、3963 或 4036 基因的异源 DNA 优选稳定转化并整合至宿主细胞的基因组中。在另一个优选实施方案中，分别含有野生型或除草剂耐受型 245、5283、2490、3963 或 4036 基因的异源 DNA 序列位于自我复制载体上。自我复制载体的例子是病毒，尤其是双生病毒。转化细胞可以再生成全株，以致该选择形式的 245、5283、2490、3963 或 4036 基因在该转基因植物中赋予除草剂耐受性。

首先将期望在转基因植物中表达的基因序列组装在表达盒中，位于能够在植物中表达的适合启动子之后。该表达盒还可以含有对于该异源 DNA 序列的表达所要求或可选择的任何其它序列。这些序列包括但不限于转录终止子、增强表达的外来序列（例如内含子）、关键序列、和旨在使该基因产物靶向特定细胞器和细胞区室的序列。然后可以容易地将这些表达盒转移到下文描述的植物转化载体上。以下描述了典型表达盒的各种成分。

用于表达盒的启动子的选择将决定该异源 DNA 序列在该 DNA 序列转化的植物中的空间和时间表达模式。所选启动子将使异源 DNA 序列在特定的细胞类型（例如叶的表皮细胞、叶肉细胞、根的皮肤细胞）中或在特定的组织或器官（例如根、叶或花）中表达，并且该选择将反映出基因产物积累的期望定位。或者，该选择的启动子可以驱动该基因在各种诱导条件下表达。启动子的强度即其启动转录的能力是不同的。根据所用的宿主细胞系统，可以使用本领域已知的许多适合启动子之任意一种。例如，对于组成型表达，可以采用 CaMV 35S 启动子、水稻肌动蛋白启动子或泛素启动子。对于可调节的表达，可以采用来自烟草或鼠耳芥属的可化学诱导的 PR-1 启动子（见例如美国专利 5,689,044）。

可以获得多种转录终止子用于表达盒中。这些转录终止子负责在该异源 DNA 序列之外的转录终止以及该 DNA 序列的正确聚腺苷酸化。适合的转录终止子是那些已知在植物中起作用的转录终止子，包括 CaMV 35S 终止子、tm1 终止子、胭脂碱合成酶终止子和豌豆 rbcS E9 终止子。这些

既可用于单子叶植物又可用于双子叶植物。

已经发现有许多序列可以增强转录单位中的基因表达，可以将这些序列与本发明的基因联合使用以增强这些基因在转基因植物中的表达。例如，各种内含子序列例如玉米 Adh1 基因的内含子已表现出可以增强表达，尤其是单子叶细胞中的表达。此外，还知道许多来源于病毒的非翻译前导序列可以增强表达，并且这些序列在双子叶细胞中尤其有效。

任选地，为了在目的农作物种中获得最佳表达可以通过改变所选基因的编码序列，以遗传改造该编码序列。修饰编码序列以实现在特定农作物种中最佳表达的方法是熟知的（见例如 Perlak 等，美国国家科学院院刊 88: 3324(1991)；和 Koziel 等，Bio/technol. 11:194 (1993)；Fennoy 和 Bailey-Serres，核酸研究 21: 5294-5300 (1993)）。考虑到植物基因中以及高等植物、绿藻和蓝细菌中密码子的选择而修饰编码序列的方法是熟知的（见 Murray 等，核酸研究 17:477-498 (1989) 中表 4；Campbell 和 Gowri 植物生理学 (Plant Physiol.) 92:1-11 (1990)）。

存在于植物中用于基因产物定向的各种机制是已知的，并且已经对控制这些机制发挥作用的序列有了较详细的特性分析。例如，基因产物导向叶绿体是由在多种蛋白质氨基末端发现的信号序列控制的，这些蛋白质在输入叶绿体的过程中被裂解形成成熟蛋白质（例如 Comai 等，生物化学杂志 263: 15104-15109 (1998)）。其它基因产物定位于其它细胞器例如线粒体和过氧化物酶体（例如 Unger 等，植物分子生物学 (Plant Molec. Biol.) 13:411-418 (1989)）。还可以对编码这些产物的 cDNA 进行操作，以实现由 DNA 序列编码的异源产物向这些细胞器的定向。此外，也已经对引起 DNA 序列所编码的产物导向其它细胞区室的序列进行了表征。氨基末端序列负责导向 ER、质外体、以及从糊粉细胞向细胞外分泌（Koehler & Ho，植物细胞 (Plant Cell) 2: 769-783 (1990)）。此外，氨基端序列联合羧基端序列负责基因产物的液泡定向（Shinshi 等，植物分子生物学 (Plant Molec. Biol.) 14:357-368 (1990)）。通过将上述的适合靶向序列与目的异源 DNA 序列融合在一起，可能指导该产物到达任何细胞器或细胞区室。

可以用于植物转化的许多转化载体是植物转化领域的普通技术人员所已知的，并且可以将本发明有关的基因联合任何这样的载体一起使用。载体的选择将取决于用于转化的优选转化技术和靶物种。对于某些靶物种，可能优选不同的抗生素或除草剂选择标记。常规用于转化的选择标记包括赋予卡那霉素以及相关抗生素抗性的 *nptII* 基因 (Messing & Vierra, 基因 (Gene) 19: 259-268 (1982); Bevan 等, 自然 304: 184-187 (1983))、赋予除草剂膦丝菌素抗性的 *bar* 基因 (White 等, 核酸研究 18: 1062 (1990); Spencer 等, 理论应用遗传学 (Theor. Appl. Genet.) 79:625-631 (1990))、赋予抗生素潮霉素抗性的 *hph* 基因 (Blochinger 和 Diggelmann, 分子细胞生物学 (Mol. Cell Biol.) 4: 2929-2931)、以及赋予氨甲蝶呤抗性的 *dhfr* 基因 (Bourouis 等, EMBO J. 2(7): 1099-1104 (1983))、和赋予草甘膦抗性的 EPSPS 基因 (美国专利 4, 940, 935 和 5, 188, 642)。

对于采用根癌农杆菌 (*Agrobacterium tumefaciens*) 的转化，有许多载体可用。这些载体典型地带有至少一个 T-DNA 边界序列，包括载体例如 pBIN19 (Bevan, 核酸研究 (1984))。适用于农杆菌转化的典型载体包括二元载体 pCIB200 和 pCIB2001，以及二元载体 pCIB10 及其潮霉素选择衍生物 (见例如美国专利 5, 639, 949)。

不使用根癌农杆菌的转化避免了在所选转化载体中对于 T-DNA 序列的要求，从而除了以上描述的含有 T-DNA 序列的那些载体之外，还可以利用缺少这些序列的载体。不依赖农杆菌的转化技术包括通过微粒轰击、原生质体摄取 (例如 PEG 和电穿孔) 和显微注射进行的转化。载体的选择很大程度上取决于对于转化物种的优选筛选方法。适用于非农杆菌转化的典型载体包括 pCIB3064、pSOG19 和 pSOG35。 (见例如美国专利 5, 639, 949)。

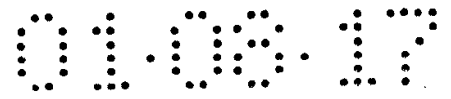
一旦目的编码序列被克隆到表达系统中，即可将其转化植物细胞。植物转化和再生的方法是本领域所熟知的。例如，Ti 质粒载体以及直接 DNA 摄取、脂质体、电穿孔、显微注射和微粒轰击已经被用于递送外源 DNA。此外，还可以利用来自农杆菌属的细菌转化植物细胞。

用于双子叶植物的转化技术是本领域所熟知的，包括基于农杆菌的技术和无需农杆菌的技术。非农杆菌技术包括原生质体或细胞直接摄取外源遗传物质。这可以通过 PEG 或电穿孔介导的摄取、微粒轰击介导的递送或显微注射来完成。在所有情况下可以采用本领域已知的标准技术将转化细胞再生成全株。

大多数单子叶植物物种的转化目前也已成为常规技术。优选的技术包括采用 PEG 或电穿孔技术直接将基因转移至原生质体中、微粒轰击导入愈伤组织、以及农杆菌介导的转化。

在另一个优选实施方案中，将编码具有粪卟啉原氧化酶活性的多肽的核苷酸序列直接转化至质体基因组中。通过同源重组将基因插入在每个植物细胞中存在的几千个拷贝的环状质体基因组中，这样与核表达基因相比该质体表达利用了巨大拷贝数的优势，从而使表达水平可以容易地超过可溶性总植物蛋白质的 10%。在一个优选的实施方案中，将所述核苷酸序列插入质体靶向载体中，并转化至期望植物宿主的质体基因组中。从而获得对于含有该核苷酸序列的质体基因组而言同质的植物，并且该植物优选能够高效表达该核苷酸序列。

质体转化技术广泛描述于例如美国专利 5,451,513、5,545,817、5,545,818 和 5,877,462，PCR 申请 WO 95/16783 和 WO 97/32977，以及 McBride 等 (1994) 美国国家科学院院刊 91, 7301-7305，所有这些文献均完整地以参考文献方式并入本文。用于质体转化的基本技术包括采用例如生物轰击 (biolistics) 或原生质体转化 (如氯化钙或 PEG 介导的转化)，将位于选择标记两侧的克隆质体 DNA 区域和所述核苷酸序列导入适合的靶组织中。该 1-1.5kb 的侧翼区域，又称靶向序列，有利于和质体基因组的同源重组，并因此使得可以对质体基因组的特定区域进行置换或修饰。最初，赋予壮观霉素和/或链霉素抗性的叶绿体 16S rRNA 和 rps12 基因中的点突变被用作转化的选择标记 (Svab, Z., Hajdukiewicz, P., 和 Maliga, P. (1990) 美国国家科学院院刊 87, 8526 - 8530; Staub, J.M. 和 Maliga, P. (1992) 植物细胞 (Plant Cell) 4, 39-45)。在这些标记之间存在的克隆位点使得可以构建用于引入外源基因的质体靶向载体



(Staub, J.M. 和 Maliga, P. (1993) EMBO J. 12, 601-606)。通过用显性选择标记置换掉该隐性 rRNA 或 r 蛋白质抗生素抗性基因, 可以获得实质性增加的转化频率, 所述显性选择标记是编码壮观霉素脱毒酶氨基糖苷-3'-腺苷转移酶的细菌 aadA 基因 (Svab, Z., 和 Maliga, P. (1993) 美国国家科学院院刊 90, 913-917)。其它对于质体转化有用的选择标记是本领域已知的, 并包括在本发明的范围内。

可以分别利用本发明野生型或改变形式的 245、5283、2490、3963 或 4036 基因, 赋予多种植物细胞除草剂耐受性, 这些植物细胞包括裸子植物、单子叶植物和双子叶植物的植物细胞。尽管可以将该基因插入属于该广泛类型的任何植物细胞中, 但在农作物植物细胞中尤其有用, 这些农作物例如水稻、小麦、大麦、黑麦、玉米、马铃薯、胡萝卜、甘薯、甜菜、菜豆、豌豆、菊苣、莴苣、卷心菜、花椰菜、嫩茎花椰菜、芜菁、萝卜、菠菜、芦笋、洋葱、大蒜、茄子、胡椒、芹菜、胡萝卜、南瓜、西葫芦、夏南瓜、黄瓜、苹果、梨、温柏、甜瓜、李子、樱桃、桃子、油桃、杏、草莓、葡萄、覆盆子、黑莓、菠萝、鳄梨、木瓜、芒果、香蕉、大豆、烟草、番茄、高粱、甘蔗。

可以将赋予植物除草剂耐受性的野生型 245、5283、2490、3963 或 4036 基因的分别高水平表达, 和/或除草剂耐受形式的 245、5283、2490、3963 或 4036 基因的分别表达, 联合其它重要的生产和质量特性, 通过本领域已知的育种方法和技术掺入植物株系中。

通过在农作物植物或可以再生出农作物植物的植物细胞培养物中直接筛选, 分别获得除草剂耐受性 245、5283、2490、3963 或 4036 等位基因之后, 可以采用传统的育种技术将这些等位基因引入商业品种中, 以开发除草剂耐受性农作物, 而无需对该等位基因进行遗传操作和用其转化该植物。

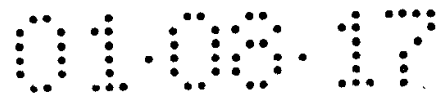
通过参考以下详细实施例, 对本发明作了进一步描述。除非另行指出, 提供这些实施例的目的仅在举例说明, 而非旨在限制。

实施例

此处所用标准重组 DNA 和分子克隆技术是本领域所熟知的, 描述于 Sambrook 等编, 分子克隆 (Molecular Cloning), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989); 和 T. J. Silhavy, M. L. Berman, 和 L. W. Enquist, 基因融合实验 (Experiments With Gene Fusions), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984); 和 Ausubel, F.M. 等, 当代分子生物学实验指南, Greene Publishing Assoc. 和 Wiley-Interscience 出版 (1987); Relter 等, 鼠耳芥属研究方法 (Methods in Arabidopsis Research), World Scientific Press (1992); 和 Schultz 等, 植物分子生物学手册 (Plant Molecular Biology Manual), Kluwer Academic Publishers (1998)。这些文献描述了用于从 T-DNA 诱变的鼠耳芥属群体中标记和克隆基因; 植物感染和转化; 筛选以鉴定幼苗突变体; 共分离分析; 和质粒拯救的所有步骤的标准技术。

实施例 1: 来自 T-DNA 诱变鼠耳芥属群体的标记幼苗致死系 # 245 的序列分析

采用质粒拯救技术分子克隆 T-DNA 诱变所产生的位于 T-DNA 插入片段一侧或两侧的鼠耳芥属基因组 DNA。通过限制性酶消化分析以该方式获得的质粒, 以便基于消化图谱对这些质粒进行分类。对于每一类质粒克隆, 确定其 DNA 序列。分析所获序列是否存在非 T-DNA 载体序列。采用 slp346for 引物 (SEQ ID NO:11) 对通过质粒拯救方法回收的质粒进行测序。引物 slp346for 提供了紧邻 T-DNA 左边界的侧翼序列的信息。通过对来自 245 突变纯合子的基因组 DNA 的 PCR, 确证质粒拯救。该 PCR 实验采用一个锚定在预测的侧翼序列中的引物和 slp346for 引物 (锚定在 T-DNA 插入片断内)。基于质粒拯救克隆的序列, 找到具有期望大小的 PCR 产物, 证实了有效拯救。将从引物 slp346for 获得序列用于核苷酸序列数据库的 BLASTx 检索 (Altschul 等 (1990), 分子生物学杂志 (J. Mol. Biol.) 215:403-410; Altschul 等 (1997) 核酸研究 25: 3389-3402)。该 BLAST 检索结果显示, 该回收的植物侧翼序列表现出与许多原核生物

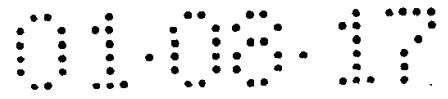


的肽释放因子 2 蛋白质的高度相似性。该 BLAST 结果指出, 该 T-DNA 插入发生在第一个鉴定的植物来源肽释放因子 2 的 ORF 中。通过采用聚合酶链式反应扩增鼠耳芥属的基因组 DNA, 分离包含肽释放因子序列相似性的 DNA 片段。将该片段用于探测 λ YES 载体中的鼠耳芥属 cDNA 文库 (Elledge 等, (1991) 美国国家科学院院刊 88: 1731-1735)。采用标准分子生物学技术分离阳性噬菌体克隆, 并对其进行特性分析。从该噬菌体切下所获 cDNA 克隆, 并确定其核苷酸序列。该 DNA 序列显示于 SEQ ID NO:1。采用 BLASTx 搜索核苷酸序列数据库, 分析推导的氨基酸序列 (Altschul 等 (1990) 分子生物学杂志 215: 403-410; Altschul 等 (1997) 核酸研究 25: 3389-3402)。该 BLAST 的搜索结果显示, 该回收的 245 cDNA 表现出与同一套原核生物的肽释放因子的序列相似性。

实施例 2: 来自 T-DNA 诱变鼠耳芥属群体的标记幼苗致死系 # 5283 的序列分析

采用质粒拯救技术分子克隆 T-DNA 诱变所产生的位于 T-DNA 插入片段一侧或两侧的鼠耳芥属基因组 DNA。通过限制性酶消化分析以该方式获得的质粒, 以便基于消化图谱对这些质粒进行分类。对于每一类质粒克隆, 确定其 DNA 序列。分析所获序列是否存在非 T-DNA 载体序列。采用 slp346for 引物 (SEQ ID NO:11) 对通过质粒拯救方法回收的质粒进行测序。引物 slp346for 提供了紧邻 T-DNA 左边界的侧翼系列的信息。通过对来自 5283 突变杂合子的基因组 DNA 的 PCR, 确证质粒拯救。该 PCR 实验采用一个锚定在预测的侧翼序列中的引物和 slp348for 引物 (SEQ ID NO:15) (锚定在 T-DNA 插入片段内)。基于质粒拯救克隆的序列, 找到具有期望大小的 PCR 产物, 证实了有效拯救。

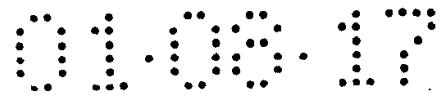
将从引物 slp346for 获得序列用于对核苷酸序列数据库进行 BLASTn 检索 (Altschul 等 (1990), 分子生物学杂志 (J. Mol. Biol.) 215:403-410; Altschul 等 (1997) 核酸研究 25: 3389-3402)。该 BLAST 检索结果显示, 该回收序列与位于鼠耳芥属第 1 号染色体中的基因组 DNA BAC T13D8 (Genbank 索取号 AC004473) 一致。设计与 BAC T13D8 序列中 #



32,964-32,987 位核苷酸反向互补的引物 LW60 (SEQ ID NO:16) (5'-aaacgcttaccatatctcttttcta-3'), 并用于确定该 T-DNA 插入片段下游的序列; 该实验鉴定了右边界的连接处。该 T-DNA 插入所发生的基因组 DNA 区域包括标注的 BAC T13D8 序列的 # 32,879 至 # 32,885 位碱基, 结果导致一个 6 碱基缺失。该插入发生在 BAC T13D8 上注释为编码与酿酒酵母 SIK1P 蛋白质相似的蛋白质 (Genbank 索取号 U20237) 的序列上游 90 个核苷酸处。通过采用聚合酶链式反应扩增鼠耳芥属的基因组 DNA, 分离包含 BAC T13D8 序列 # 33,025 位碱基至 # 34,338 位碱基的 DNA 片段。将该片段用于探测 IYES 载体中的鼠耳芥属 cDNA 文库 (Elledge 等, (1991) 美国国家科学院院刊 88: 1731-1735)。采用标准分子生物学技术分离阳性噬菌体克隆, 并对其进行特性分析。从该噬菌体切下所获 cDNA 克隆, 并确定其核苷酸序列。鉴定了一个全长克隆。采用 tBLASTn 检索核苷酸序列数据库, 分析推导的氨基酸序列 (Altschul 等 (1990) 分子生物学杂志 215: 403-410; Altschul 等 (1997) 核酸研究 25: 3389-3402)。该 BLAST 的检索结果显示, 该回收的 5283 cDNA 序列来源于位于鼠耳芥属第 1 号染色体的相同基因组序列 BAC T13D8。除了以下不同之外, 该 cDNA 序列的内含子/外显子边界与就鼠耳芥属 SIK1P 同系物 (Genbank 索取号 AC004473) 所预测的相同。对于 5283 cDNA 起始密码子由 # 32975 位碱基 - # 32977 位碱基编码, 之后紧接一个从 # 32978 至 # 33199 位碱基的内含子。

实施例 3: 来自 T-DNA 诱变鼠耳芥属群体的标记幼苗致死系 # 2490 的序列分析

采用质粒拯救技术分子克隆 T-DNA 诱变所产生的位于 T-DNA 插入片段一侧或两侧的鼠耳芥属基因组 DNA。通过限制性酶消化分析以该方式获得的质粒, 以便基于消化图谱对这些质粒进行分类。对于每一类质粒克隆, 确定其 DNA 序列。分析所获序列是否存在非 T-DNA 载体序列。采用 SLP346for 引物 (5'GCGGACATCTACATTTTTGA3':SEQ ID NO:11), 对通过质粒拯救方法回收的质粒进行测序。引物 SLP346for 提供了紧邻 T-DNA



左边界的侧翼序列的信息。作为含有 T-DNA 左边界的质粒，回收该 T-DNA 插入片段两端的克隆。通过 Southern 印迹分析比较来自植物 2490 突变杂合子的基因组 DNA 和来自植物野生型 2490 基因纯合子的基因组 DNA，验证质粒拯救。从使用 SLP369 (5'CAGACCA CAATACCTTCAAAAATA3': SEQ ID NO: 22) 和 SLP370 (5'CCATTGTGTCTCCCTCC CGCTGTT3': SEQ ID NO:23) 引物产生的 PCR 产物，制备 Southern 印迹的探针。在该 2490 杂合子中找到另一个 BamH1 片段，验证了有效拯救。

将从以上克隆获得的序列用于对核苷酸序列数据库进行 BLASTn 检索 (Altschul 等 (1990), 分子生物学杂志 (J. Mol. Biol.) 215:403-410; Altschul 等 (1997) 核酸研究 25: 3389-3402)。该检索结果显示，该回收序列与鼠耳芥属第 5 号染色体 P1 克隆 MTG13 (Genbank # AB008270) 中的基因组 DNA 一致。当将该插入事件发生处的基因组 DNA 区域用于 Genbank EST 数据库的 BLASTn 检索时，鉴定了来源于两个 EST 末端的 4 个序列: 144K24 (144K24 T7 Genbank# T76608 和 144K24XP Genbank # AA404903) 和 GBGF153 (5'端 Genbank#F15182 和 3'端 Genbank#F15181)。确定了 144K24 EST 的完整序列，该序列编码 2490 基因的完整可读框 (ORF)。该 EST 的 BLAST 分析指出，该 2490 蛋白质与欧洲油菜 (*Brassica napus*) Toc36 蛋白质 (Genbank #X79091; Ko 等 (1995) 生物化学杂志 270: 28601-28608; Wu 等 (1994) 生物化学杂志 269: 32264-32271; Pang 等 (1997) 生物化学杂志 272: 25623-25627) 有序列相似性。该 Toc36 蛋白质还被称作 bce44B、Com44 和 Cim44。由于含有该 2490 ORF 的基因组 DNA 直到目前仍没有被正确地标释，所以本发明人第一个提供了该 2490 基因的正确 ORF 和序列相似性的实验资料。

实施例 4: 来自 T-DNA 诱变鼠耳芥属群体的标记幼苗致死系 # 3963 的序列分析

采用质粒拯救技术分子克隆 T-DNA 诱变所产生的位于 T-DNA 插入片段一侧或两侧的鼠耳芥属基因组 DNA。通过限制性酶消化分析以该方式获得的质粒，以便基于消化图谱对这些质粒进行分类。对于每一类质粒克

隆，确定其 DNA 序列。分析所获序列是否存在非 T-DNA 载体序列。采用-21 引物（5'TGTA AACGACGGCCAGT 3' : SEQ ID NO:25），对通过质粒拯救方法回收的质粒进行测序。引物-21 提供了紧邻 T-DNA 右边界的侧翼序列的信息。通过从 3963 突变杂合子 PCR 扩增基因组 DNA，验证质粒拯救。该 PCR 实验采用一个锚定在预测的侧翼序列中的引物和-21 引物（锚定在该 T-DNA 插入片段中）。基于该质粒拯救克隆的序列，找到具有期望大小的 PCR 产物，证实了有效拯救。将从引物-21 获得的序列用于对核苷酸序列数据库进行 BLASTn 检索(Altschul 等(1990), 分子生物学杂志(J. Mol. Biol.) 215:403-410; Altschul 等(1997) 核酸研究 25: 3389-3402)。该 BLAST 检索结果显示，该回收的植物侧翼序列与第 5 号染色体上 P1 克隆 MDK4 (Genbank 索取号 AB010695) 中的基因组序列 100% 一致。该 T-DNA 插入片段发生在标注的 P1 克隆 MDK4 序列的 # 36342 位碱基处，在鉴定为 MDK4.6 的基因中。该回收侧翼序列的 tBLASTX 分析显示出与酿酒酵母 DNA 修复蛋白质 Mre11p (Genbank 索取号 U60829) 的序列相似性。通过采用聚合酶链式反应扩增鼠耳芥属基因组 DNA，分离编码该鼠耳芥属 3963 蛋白质部分的片段。该片段被用于探测 λ YES 载体中的鼠耳芥属 cDNA 文库 (Elledge 等(1991) 美国国家科学院院刊 88: 1731-1735)。采用标准分子生物学技术分离阳性噬菌体克隆，并对其进行特性分析。从该噬菌体切下所获 cDNA 克隆，并确定其核苷酸序列。鉴定了一个 cDNA 克隆。该 cDNA 序列显示于 SEQ ID NO:7。采用 BLASTx 检索核苷酸序列数据库，分析推导的氨基酸序列(Altschul 等(1990), 分子生物学杂志 215: 403-410; Altschul 等(1997) 核酸研究 25: 3389-3402)。该 BLAST 结果显示，该回收的 3963 cDNA 表现出与许多 DNA 修复蛋白质的序列相似性，这些蛋白质包括粟酒裂殖酵母的 Rad32p (Genbank 索取号 Q09683); 人的 hMre11 (Genbank 索取号 U37359); 和酿酒酵母的 Mre11p (Genbank 索取号 U60829)。由于含有该 3963 开发阅读框(ORF)的基因组 DNA 就外显子/内含子边界而言在现有技术中未有正确标注，所以本发明人第一个提供了 3963 基因正确 ORF 的实验资料。现有技术指出这些外显子/内含子边界: 35662-35817、36015-36172、36315-36405、36528-36647、36728

- 36796、36865 - 36956、37045 - 37147、37247 - 37354、37476 - 37538、37785 - 37862、38060 - 38122、38211 - 38271、38753 - 38835、38979 - 39092、39468 - 39766、39879 - 40002、40161 - 40370。相应于本文公开的部分 cDNA 的外显子/内含子边界是：位置不详的 5' 末端（第一个知道的碱基在第 36147 位）、36147 - 36172、36315 - 36405、36528 - 36647、36728 - 36796、36865 - 36956、37045 - 37147、37247 - 37354、37476 - 37538、37610 - 37681、37785 - 39092、39212 - 39290、39377 - 39445、39532 - 39776、39879 - 40002、40161 - 40363、40478 - 40508（终止子开始于 40509）。

实施例 5: 来自 T-DNA 诱变鼠耳芥属群体的标记幼苗致死系 # 4036 的序列分析

采用质粒拯救技术分子克隆 T-DNA 诱变所产生的位于 T-DNA 插入片段一侧或两侧的鼠耳芥属基因组 DNA。通过限制性酶消化分析以该方式获得的质粒，以便基于消化图谱对这些质粒进行分类。对于每一类质粒克隆，确定其 DNA 序列。分析所获序列是否存在非 T-DNA 载体序列。采用 slp346 引物（5'GCGGACATCTACATTTTTGA3'；SEQ ID NO:11），对通过质粒拯救方法回收的质粒进行测序。引物 slp346 提供了紧邻 T-DNA 左边界的侧翼序列的信息。通过 PCR 扩增来自 4036 插入突变杂合子的模板基因组 DNA，验证质粒拯救。该实验采用一个锚定在预测的侧翼序列中的引物和 slp328 引物（5'ACCTTAGGCGACTTTTGAAC3'；SEQ ID NO:15；锚定在该 T-DNA 插入片段中）。基于该质粒拯救克隆的序列，找到具有期望大小的 PCR 产物，证实了有效拯救。

将从上述克隆获得的序列用于对核苷酸序列数据库进行 BLASTn 检索（Altschul 等 (1990)，分子生物学杂志 (J. Mol. Biol.) 215:403-410；Altschul 等 (1997) 核酸研究 25: 3389-3402)。该 BLAST 检索结果显示，该植物侧翼序列与来自鼠耳芥属第 5 号染色体的 P1 MQB2 公开基因组序列（Genbank 索取号 AB003053）100% 一致。该 T-DNA 插入发生在标注的 P1 克隆的 # 31, 380 位碱基处，并中断了鉴定为 MQB2.6 的基因。

由该中断的可读框(ORF)编码的蛋白质表现出与许多生物的 1-脱氧-D-木酮糖 5-磷酸还原异构酶的相似性, 这些生物包括集胞蓝细菌属 (*Synechocystis* sp.) (SWISS-PROTQ55663)、枯草芽孢杆菌 (SWISS-PROT031753) 和大肠杆菌 (SWISS-PROT P45568) (Takahashi 等 (1998) 美国国家科学院院刊 95: 9879 - 9884)。用 Web GeneMark 软件 (Borodovsky, M. 和 Mcininch J. (1993) 计算机和化学 (Computers & Chemistry) 17:123-133), 对含有该 ORF 的基因组区域进行重新标注。然后针对该预测 ORF 的 5' 和 3' 末端设计引物, 并采用来自 pFL61 鼠耳芥属 cDNA 文库的 DNA (Minet 等 (1992) 植物杂志 (Plant J.) 2:417-422) 作为模板进行 PCR。TA 连接和克隆所获 PCR 产物 (Original TA 克隆试剂盒, Invitrogen), 并对其进行测序。由于含有该 4036 ORF 的基因组 DNA 就外显子/内含子边界而言在现有技术中未有正确标注, 所以本发明人第一个提供了 4036 基因的正确 ORF 的实验资料。现有技术指出这些外显子/内含子边界: 33490..33356、31293..31207、30971..30846、30780..30718、30622..30473、30345..30288、30194..30083、29996..29892、29805..29684、29394..29248、29162..28997。在本发明的序列中, 第 31928 位碱基标记为该 cDNA 起始密码子的第一个碱基, 第 28996 位碱基标记为该 cDNA 终止密码子的第一个碱基。含有该起始密码子的外显子的 3' 末端是 31836, 而含有该终止密码子的外显子的 5' 末端是 29161。本文公开的该 cDNA 的内部外显子/内含子边界是: 31640..31448、31294..31202、30965..30843、30777..30722、30636..30473、30355..30287、30193..30082、29995..29891、29804..29684、29394..29247。

实施例 6a 在大肠杆菌中表达重组 245 蛋白质

将相应于 cDNA 克隆 SEQ ID NO:1 的蛋白质编码区亚克隆至前面所描述的表达载体中, 并采用厂家的条件转化大肠杆菌。具体的实例包括质粒例如 pBluescript (Stratagene, La Jolla, CA)、pFLAG (International Biotechnologies, Inc., New Haven, CT)、和 pTrcHis

(Invitrogen, La Jolla, CA)。培养大肠杆菌，并验证 245 活性的表达。采用标准技术分离赋予 245 活性的蛋白质。

实施例 6b 在大肠杆菌中表达重组 5283 蛋白质

将相应于 cDNA 克隆 SEQ ID NO:3 的蛋白质编码区亚克隆至前面所描述的表达载体中，并采用厂家的条件转化大肠杆菌。具体的实例包括质粒例如 pBluescript (Stratagene, La Jolla, CA)、pFLAG (International Biotechnologies, Inc., New Haven, CT)、和 pTrcHis (Invitrogen, La Jolla, CA)。培养大肠杆菌，并验证 5283 活性的表达。采用标准技术分离赋予 5283 活性的蛋白质。

实施例 6c 在大肠杆菌中表达重组 2490 蛋白质

将相应于 cDNA 克隆 SEQ ID NO:5 的蛋白质编码区亚克隆至前面所描述的表达载体中，并采用厂家的条件转化大肠杆菌。具体的实例包括质粒例如 pBluescript (Stratagene, La Jolla, CA)、pFLAG (International Biotechnologies, Inc., New Haven, CT)、和 pTrcHis (Invitrogen, La Jolla, CA)。培养大肠杆菌，并验证 2490 活性的表达。采用标准技术分离赋予 2490 活性的蛋白质。

实施例 6d 在大肠杆菌中表达重组 3963 蛋白质

将相应于 cDNA 克隆 SEQ ID NO:7 的蛋白质编码区亚克隆至前面所描述的表达载体中，并采用厂家的条件转化大肠杆菌。具体的实例包括质粒例如 pBluescript (Stratagene, La Jolla, CA)、pFLAG (International Biotechnologies, Inc., New Haven, CT)、和 pTrcHis (Invitrogen, La Jolla, CA)。培养大肠杆菌，并验证 3963 活性的表达。采用标准技术分离赋予 3963 活性的蛋白质。

实施例 6e 在大肠杆菌中表达重组 4036 蛋白质

将相应于 cDNA 克隆 SEQ ID NO:9 的蛋白质编码区亚克隆至前面所描

述的表达载体中，并采用厂家的条件转化大肠杆菌。具体的实例包括质粒例如 pBluescript (Stratagene, La Jolla, CA)、pFLAG (International Biotechnologies, Inc., New Haven, CT)、和 pTrcHis (Invitrogen, La Jolla, CA)。培养大肠杆菌，并验证 4036 活性的表达。采用标准技术分离赋予 4036 活性的蛋白质。

实施例 7: 通过 DNA 改组体外重组 245、5283、2490、3963 或 4036 基因

通过 PCR 分别扩增 SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7 或 SEQ ID NO:9 的核苷酸序列。基本上按已描述的方法 (Stemmer 等 (1994) PNAS 91: 10747-10751) 通过 DNaseI 处理消化该所获得的 DNA 片段，并从该反应混合物中除去 PCR 引物。在没有引物的情况下进行 PCR 反应，之后在有引物的情况下进行 PCR 反应，两者均按已描述的方法进行 (Stemmer 等 (1994) PNAS 91: 10747-10751)。将所获得的 DNA 片段克隆至用于大肠杆菌的 pTRC99a (Pharmacia, 产品号: 27-5007-01) 中，或克隆至用于酵母的 pESC 载体中 (Stratagene 产品目录) 中；然后采用 Biorad Gene Pulser 以及厂家的条件，通过电穿孔转化分别缺少 245、5283、2490、3963 或 4036 活性的细菌或酵母株。将该转化细菌或酵母培养在含有抑制性浓度的 245、5283、2490、3963 或 4036 活性抑制剂的培养基上，选择存在该抑制剂时生长的菌落。挑取在有正常抑制性浓度的抑制剂时生长的菌落，并通过重复再划线进行纯化。纯化它们的质粒，然后确定通过该测试的质粒 cDNA 插入片段的 DNA 序列。

在相似的反应中，将分别含有鼠耳芥 245、5283、2490、3963 或 4036 基因、编码蛋白质的 PCR 扩增 DNA 片段，和分别含有来自大肠杆菌的 245、5283、2490、3963 或 4036 基因的 PCR 扩增 DNA 片段在体外重组，并按以上所述回收获得的对抑制剂具有提高的耐受性的变异体。

实施例 8a: 通过交错延伸方法体外重组 245 基因

将编码 245 蛋白质的鼠耳芥 245 基因和大肠杆菌 245 基因各克隆至

pBluescript 载体的多位点接头中。采用“反向引物”和“M13 -20 引物” (Stratagene 产品目录), 基本按已描述的方法 (Zhao 等 (1998) 自然生物技术 16: 258-261), 进行 PCR 反应。用适合的限制性酶消化扩增的 PCR 片段, 并将其克隆至 pTRC99a 中, 按实施例 7 所描述的方法筛选突变的 245 基因。

实施例 8b: 通过交错延伸方法体外重组 5283 基因

将编码 5283 蛋白质的鼠耳芥 5283 基因和大肠杆菌 5283 基因各克隆至 pBluescript 载体的多位点接头中。采用“反向引物”和“M13 -20 引物” (Stratagene 产品目录), 基本按已描述的方法 (Zhao 等 (1998) 自然生物技术 16: 258-261), 进行 PCR 反应。用适合的限制性酶消化扩增的 PCR 片段, 并将其克隆至 pTRC99a 中, 按实施例 7 所描述的方法筛选突变的 5283 基因。

实施例 8c: 通过交错延伸方法体外重组 2490 基因

将编码 2490 蛋白质的鼠耳芥 2490 基因和大肠杆菌 2490 基因各克隆至 pBluescript 载体的多位点接头中。采用“反向引物”和“M13 -20 引物” (Stratagene 产品目录), 基本按已描述的方法 (Zhao 等 (1998) 自然生物技术 16: 258-261), 进行 PCR 反应。用适合的限制性酶消化扩增的 PCR 片段, 并将其克隆至 pTRC99a 中, 按实施例 7 所描述的方法筛选突变的 2490 基因。

实施例 8d: 通过交错延伸方法体外重组 3963 基因

将编码 3963 蛋白质的鼠耳芥 3963 基因和大肠杆菌 3963 基因各克隆至 pBluescript 载体的多位点接头中。采用“反向引物”和“M13 -20 引物” (Stratagene 产品目录), 基本按已描述的方法 (Zhao 等 (1998) 自然生物技术 16: 258-261), 进行 PCR 反应。用适合的限制性酶消化扩增的 PCR 片段, 并将其克隆至 pTRC99a 中, 按实施例 7 所描述的方法筛选突变的 3963 基因。

实施例 8e: 通过交错延伸方法体外重组 4036 基因

将编码 4036 蛋白质的鼠耳芥 4036 基因和大肠杆菌 4036 基因各克隆至 pBluescript 载体的多位点接头中。采用“反向引物”和“M13 -20 引物”(Stratagene 产品目录),基本按已描述的方法(Zhao 等 (1998) 自然生物技术 16: 258-261),进行 PCR 反应。用适合的限制性酶消化扩增的 PCR 片段,并将其克隆至 pTRC99a 中,按实施例 7 所描述的方法筛选突变的 4036 基因。

实施例 9: 体外结合分析

例如分别根据实施例 6a、6b、6c、6d 或 6e,获得重组 245、5283、2490、3963 或 4036 蛋白质。采用本领域熟知的技术,将该蛋白质固定在适于配体结合分析的芯片上。根据本领域所熟知的方法将固定在该芯片上的蛋白质暴露于溶液中的样品化合物。当该样品化合物与该固定的蛋白质接触后,实施能够检测到蛋白质-配体相互作用的测量方法。这些测量方法的例子有以上描述的 SELDI、biacore 和 FCS。以这种方式可以容易地发现与该蛋白质结合的化合物,并对其作进一步的特性分析。

以上公开的实施方案是举例说明用的。本发明的公开将使本领域的技术人员拥有许多本发明的变化。所有这些明显和可预见的变化均包含在所附的权利要求之内。

序列表

<110> Novartis AG

<120> 除草剂的靶基因和方法

<130> 组合的除草剂靶基因

<140>

<141>

<160> 29

<170> PatentIn Ver. 2.1

<210> 1

<211> 1119

<212> DNA

<213> 鼠耳芥

<220>

<221> CDS

<222> (1)..(1119)

<400> 1

```

atg gat gac atg gac acc gtc tac aag caa ttg gga ttg ttt tca cta      48
Met Asp Asp Met Asp Thr Val Tyr Lys Gln Leu Gly Leu Phe Ser Leu
  1             5             10             15

aag aag aag att aaa gat gtt gtt ctt aag gct gag atg ttt gca ccg      96
Lys Lys Lys Ile Lys Asp Val Val Leu Lys Ala Glu Met Phe Ala Pro
             20             25             30

gat gct ctt gag ctt gaa gaa gag cag tgg ata aag caa gaa gaa aca     144
Asp Ala Leu Glu Leu Glu Glu Glu Gln Trp Ile Lys Gln Glu Glu Thr
             35             40             45

atg cgt tac ttt gat tta tgg gat gat ccc gct aaa tct gat gag att     192
Met Arg Tyr Phe Asp Leu Trp Asp Asp Pro Ala Lys Ser Asp Glu Ile
             50             55             60

ctt ctc aaa tta gct gat cga gct aaa gca gtc gat tcc ctc aaa gac     240
Leu Leu Lys Leu Ala Asp Arg Ala Lys Ala Val Asp Ser Leu Lys Asp
             65             70             75             80

ctc aaa tac aag gct gaa gaa gct aag ctg atc ata caa ttg ggt gag     288
Leu Lys Tyr Lys Ala Glu Glu Ala Lys Leu Ile Ile Gln Leu Gly Glu
             85             90             95

atg gat gct ata gat tac agt ctc ttt gag caa gcc tat gat tca tca     336
Met Asp Ala Ile Asp Tyr Ser Leu Phe Glu Gln Ala Tyr Asp Ser Ser
             100            105            110

ctc gat gta agt aga tcg ttg cat cac tat gag atg tct aag ctt ctt     384
Leu Asp Val Ser Arg Ser Leu His His Tyr Glu Met Ser Lys Leu Leu

```


115	120	125	
agg gat caa tat gac gct gaa ggc gct tgt atg att atc aaa tct gga			432
Arg Asp Gln Tyr Asp Ala Glu Gly Ala Cys Met Ile Ile Lys Ser Gly			
130	135	140	
tct cca ggc gca aaa tct cag ata tgg aca gag caa gtt gta agt atg			480
Ser Pro Gly Ala Lys Ser Gln Ile Trp Thr Glu Gln Val Val Ser Met			
145	150	155	160
tat atc aaa tgg gca gaa agg cta ggc caa aac gcg cgg gtg gct gag			528
Tyr Ile Lys Trp Ala Glu Arg Leu Gly Gln Asn Ala Arg Val Ala Glu			
	165	170	175
aaa tgt agt tta ttg agt aat aaa agt ggc gta agt tca gcc acg ata			576
Lys Cys Ser Leu Leu Ser Asn Lys Ser Gly Val Ser Ser Ala Thr Ile			
	180	185	190
gag ttt gaa ttc gag ttt gct tat ggt tat ctc tta ggt gag cga ggt			624
Glu Phe Glu Phe Glu Phe Ala Tyr Gly Tyr Leu Leu Gly Glu Arg Gly			
	195	200	205
gtg cac cgc ctt atc ata agt tcc act tct aat gag gaa tgt tca gcg			672
Val His Arg Leu Ile Ile Ser Ser Thr Ser Asn Glu Glu Cys Ser Ala			
	210	215	220
act gtt gat atc ata cca cta ttc ttg aga gca tct cct gat ttt gaa			720
Thr Val Asp Ile Ile Pro Leu Phe Leu Arg Ala Ser Pro Asp Phe Glu			
225	230	235	240
gta aag gaa ggt gat ttg att gta tcg tat cct gca aaa gag gat cac			768
Val Lys Glu Gly Asp Leu Ile Val Ser Tyr Pro Ala Lys Glu Asp His			
	245	250	255
aaa ata gct gag aat atg gtt tgt atc cac cat att ccg agt gga gta			816
Lys Ile Ala Glu Asn Met Val Cys Ile His His Ile Pro Ser Gly Val			
	260	265	270
aca cta caa tct tca gga gaa aga aac cgg ttt gca aac agg atc aaa			864
Thr Leu Gln Ser Ser Gly Glu Arg Asn Arg Phe Ala Asn Arg Ile Lys			
	275	280	285
gct cta aac cgg ttg aag gcg aag cta ctt gtg ata gca aaa gag caa			912
Ala Leu Asn Arg Leu Lys Ala Lys Leu Leu Val Ile Ala Lys Glu Gln			
	290	295	300
aag gtt tcg gat gta aat aaa atc gac agc aag aac att ttg gaa ccg			960
Lys Val Ser Asp Val Asn Lys Ile Asp Ser Lys Asn Ile Leu Glu Pro			
305	310	315	320
cgg gaa gaa acc agg agt tat gtc tct aag ggt cac aag atg gtg gtt			1008
Arg Glu Glu Thr Arg Ser Tyr Val Ser Lys Gly His Lys Met Val Val			
	325	330	335
gat aga aaa acc ggt tta gag att ctg gac ctg aaa tcg gtc ttg gat			1056

Asp Arg Lys Thr Gly Leu Glu Ile Leu Asp Leu Lys Ser Val Leu Asp
 340 345 350

gga aac att gga cca ctc ctt gga gct cat att agc atg aga aga tca 1104
 Gly Asn Ile Gly Pro Leu Leu Gly Ala His Ile Ser Met Arg Arg Ser
 355 360 365

att gat gcg att tag 1119
 Ile Asp Ala Ile
 370

- <210> 2
- <211> 372
- <212> PRT
- <213> 鼠耳芥

<400> 2
 Met Asp Asp Met Asp Thr Val Tyr Lys Gln Leu Gly Leu Phe Ser Leu
 1 5 10 15
 Lys Lys Lys Ile Lys Asp Val Val Leu Lys Ala Glu Met Phe Ala Pro
 20 25 30
 Asp Ala Leu Glu Leu Glu Glu Glu Gln Trp Ile Lys Gln Glu Glu Thr
 35 40 45
 Met Arg Tyr Phe Asp Leu Trp Asp Asp Pro Ala Lys Ser Asp Glu Ile
 50 55 60
 Leu Leu Lys Leu Ala Asp Arg Ala Lys Ala Val Asp Ser Leu Lys Asp
 65 70 75 80
 Leu Lys Tyr Lys Ala Glu Glu Ala Lys Leu Ile Ile Gln Leu Gly Glu
 85 90 95
 Met Asp Ala Ile Asp Tyr Ser Leu Phe Glu Gln Ala Tyr Asp Ser Ser
 100 105 110
 Leu Asp Val Ser Arg Ser Leu His His Tyr Glu Met Ser Lys Leu Leu
 115 120 125
 Arg Asp Gln Tyr Asp Ala Glu Gly Ala Cys Met Ile Ile Lys Ser Gly
 130 135 140
 Ser Pro Gly Ala Lys Ser Gln Ile Trp Thr Glu Gln Val Val Ser Met
 145 150 155 160
 Tyr Ile Lys Trp Ala Glu Arg Leu Gly Gln Asn Ala Arg Val Ala Glu
 165 170 175
 Lys Cys Ser Leu Leu Ser Asn Lys Ser Gly Val Ser Ser Ala Thr Ile
 180 185 190
 Glu Phe Glu Phe Glu Phe Ala Tyr Gly Tyr Leu Leu Gly Glu Arg Gly
 195 200 205
 Val His Arg Leu Ile Ile Ser Ser Thr Ser Asn Glu Glu Cys Ser Ala
 210 215 220
 Thr Val Asp Ile Ile Pro Leu Phe Leu Arg Ala Ser Pro Asp Phe Glu
 225 230 235 240
 Val Lys Glu Gly Asp Leu Ile Val Ser Tyr Pro Ala Lys Glu Asp His
 245 250 255
 Lys Ile Ala Glu Asn Met Val Cys Ile His His Ile Pro Ser Gly Val
 260 265 270
 Thr Leu Gln Ser Ser Gly Glu Arg Asn Arg Phe Ala Asn Arg Ile Lys
 275 280 285
 Ala Leu Asn Arg Leu Lys Ala Lys Leu Leu Val Ile Ala Lys Glu Gln

```

      290              295              300
Lys Val Ser Asp Val Asn Lys Ile Asp Ser Lys Asn Ile Leu Glu Pro
305              310              315              320
Arg Glu Glu Thr Arg Ser Tyr Val Ser Lys Gly His Lys Met Val Val
      325              330              335
Asp Arg Lys Thr Gly Leu Glu Ile Leu Asp Leu Lys Ser Val Leu Asp
      340              345              350
Gly Asn Ile Gly Pro Leu Leu Gly Ala His Ile Ser Met Arg Arg Ser
      355              360              365
Ile Asp Ala Ile
      370

```

```

<210> 3
<211> 1458
<212> DNA
<213> 鼠耳芥

```

```

<220>
<221> CDS
<222> (1)..(1458)

```

```

<400> 3
atg gca act ctt gaa gat tct ttc ctt gct gat ttg gac gag tta tct 48
Met Ala Thr Leu Glu Asp Ser Phe Leu Ala Asp Leu Asp Glu Leu Ser
  1              5              10              15

gac aat gaa gca gaa ttg gac gag aat gat ggt gat gtt gga aag gaa 96
Asp Asn Glu Ala Glu Leu Asp Glu Asn Asp Gly Asp Val Gly Lys Glu
      20              25              30

gaa gaa gat gtt gat atg gat atg gct gat tta gag aca ctt aac tat 144
Glu Glu Asp Val Asp Met Asp Met Ala Asp Leu Glu Thr Leu Asn Tyr
      35              40              45

gat gat ctc gat aat gtt tct aag ctg cag aag agt cag aga tat gct 192
Asp Asp Leu Asp Asn Val Ser Lys Leu Gln Lys Ser Gln Arg Tyr Ala
      50              55              60

gat att atg cat aaa gta gag gag gct ctt ggg aaa gat tct gat gga 240
Asp Ile Met His Lys Val Glu Glu Ala Leu Gly Lys Asp Ser Asp Gly
      65              70              75              80

gct gag aaa gga act gtc ttg gaa gat gat cct gag tat aag ctt att 288
Ala Glu Lys Gly Thr Val Leu Glu Asp Asp Pro Glu Tyr Lys Leu Ile
      85              90              95

gtg gat tgt aat cag ctt tcg gtc gat att gag aat gaa atc gtt att 336
Val Asp Cys Asn Gln Leu Ser Val Asp Ile Glu Asn Glu Ile Val Ile
      100             105             110

gtc cac aac ttt atc aaa gac aag tac aag ctt aag ttt caa gag ctt 384
Val His Asn Phe Ile Lys Asp Lys Tyr Lys Leu Lys Phe Gln Glu Leu
      115             120             125

```

gag tcg ttg gtt cat cac cct att gac tat gca tgt gtt gtg aag aag 432
 Glu Ser Leu Val His His Pro Ile Asp Tyr Ala Cys Val Val Lys Lys
 130 135 140

att ggg aat gag acg gat ttg gct ctt gtt gat ctc gct gac ctt ctt 480
 Ile Gly Asn Glu Thr Asp Leu Ala Leu Val Asp Leu Ala Asp Leu Leu
 145 150 155 160

cct tca gct att atc atg gtt gtt tca gtt act gct tta act acg aaa 528
 Pro Ser Ala Ile Ile Met Val Val Ser Val Thr Ala Leu Thr Thr Lys
 165 170 175

ggg agt gca ctg cca gag gat gtt ttg caa aag gtg tta gag gct tgt 576
 Gly Ser Ala Leu Pro Glu Asp Val Leu Gln Lys Val Leu Glu Ala Cys
 180 185 190

gat cgg gct tta gat ctt gat tcc gca agg aag aag gtc ctt gag ttt 624
 Asp Arg Ala Leu Asp Leu Asp Ser Ala Arg Lys Lys Val Leu Glu Phe
 195 200 205

gtt gaa agt aag atg gga tct att gca cct aat ctt tct gct att gtt 672
 Val Glu Ser Lys Met Gly Ser Ile Ala Pro Asn Leu Ser Ala Ile Val
 210 215 220

ggg agt gct gtt gca gcc aaa ctc atg ggg act gct gga ggt ttg tca 720
 Gly Ser Ala Val Ala Ala Lys Leu Met Gly Thr Ala Gly Gly Leu Ser
 225 230 235 240

gca ctt gct aaa atg cct gcg tgt aat gtt caa gtt ctt ggc cac aag 768
 Ala Leu Ala Lys Met Pro Ala Cys Asn Val Gln Val Leu Gly His Lys
 245 250 255

agg aag aac ctt gct ggg ttt tct tct gca acg tct cag tcc cgt gtg 816
 Arg Lys Asn Leu Ala Gly Phe Ser Ser Ala Thr Ser Gln Ser Arg Val
 260 265 270

ggt tat ctg gag cag aca gag att tac caa agc acg cct cct gga ctt 864
 Gly Tyr Leu Glu Gln Thr Glu Ile Tyr Gln Ser Thr Pro Pro Gly Leu
 275 280 285

cag gct cgc gct ggc agg ctc gtg gct gca aaa tca act ttg gca gca 912
 Gln Ala Arg Ala Gly Arg Leu Val Ala Ala Lys Ser Thr Leu Ala Ala
 290 295 300

aga gtt gat gct act aga ggg gat ccg tta ggg ata agt gga aaa gct 960
 Arg Val Asp Ala Thr Arg Gly Asp Pro Leu Gly Ile Ser Gly Lys Ala
 305 310 315 320

ttc agg gag gag atc cgt aag aag att gag aaa tgg caa gaa cct cct 1008
 Phe Arg Glu Glu Ile Arg Lys Lys Ile Glu Lys Trp Gln Glu Pro Pro
 325 330 335

cct gca aga cag cct aag cca ctt cct gtt cct gat tct gaa ccg aag 1056
 Pro Ala Arg Gln Pro Lys Pro Leu Pro Val Pro Asp Ser Glu Pro Lys

340	345	350	
aaa aga agg ggt ggt cgc cgt cta aga aaa atg aaa gaa agg tat caa			1104
Lys Arg Arg Gly Gly Arg Arg Leu Arg Lys Met Lys Glu Arg Tyr Gln			
355	360	365	
gta aca gat atg agg aag ctg gcc aac aga atg gcg ttt ggt aca cct			1152
Val Thr Asp Met Arg Lys Leu Ala Asn Arg Met Ala Phe Gly Thr Pro			
370	375	380	
gaa gag agc tcc ctc ggt gat gga cta gga gaa ggt tat gga atg ctt			1200
Glu Glu Ser Ser Leu Gly Asp Gly Leu Gly Glu Gly Tyr Gly Met Leu			
385	390	395	400
ggc cag gca gga agc aac agg ctg cga gta tcc agt gtt ccg agc aag			1248
Gly Gln Ala Gly Ser Asn Arg Leu Arg Val Ser Ser Val Pro Ser Lys			
405	410	415	
ctt aag att aat gct aag gtc gcc aaa aag ctt aaa gaa agg cag tat			1296
Leu Lys Ile Asn Ala Lys Val Ala Lys Lys Leu Lys Glu Arg Gln Tyr			
420	425	430	
gcg ggt ggt gcg act acc tct ggt ttg aca tcg agc ctg gct ttc act			1344
Ala Gly Gly Ala Thr Thr Ser Gly Leu Thr Ser Ser Leu Ala Phe Thr			
435	440	445	
cct gtg cag gga ata gag ttg tgc aat cct cag cag gct tta gga tta			1392
Pro Val Gln Gly Ile Glu Leu Cys Asn Pro Gln Gln Ala Leu Gly Leu			
450	455	460	
gga agt ggg act caa agc act tac ttc tca gag tca gga acc ttc tcg			1440
Gly Ser Gly Thr Gln Ser Thr Tyr Phe Ser Glu Ser Gly Thr Phe Ser			
465	470	475	480
aag ctg aag aag atc taa			1458
Lys Leu Lys Lys Ile			
485			

<210> 4
 <211> 485
 <212> PRT
 <213> 鼠耳芥

<400> 4
 Met Ala Thr Leu Glu Asp Ser Phe Leu Ala Asp Leu Asp Glu Leu Ser
 1 5 10 15
 Asp Asn Glu Ala Glu Leu Asp Glu Asn Asp Gly Asp Val Gly Lys Glu
 20 25 30
 Glu Glu Asp Val Asp Met Asp Met Ala Asp Leu Glu Thr Leu Asn Tyr
 35 40 45
 Asp Asp Leu Asp Asn Val Ser Lys Leu Gln Lys Ser Gln Arg Tyr Ala
 50 55 60
 Asp Ile Met His Lys Val Glu Glu Ala Leu Gly Lys Asp Ser Asp Gly
 65 70 75 80

Ala Glu Lys Gly Thr Val Leu Glu Asp Asp Pro Glu Tyr Lys Leu Ile
85 90 95
Val Asp Cys Asn Gln Leu Ser Val Asp Ile Glu Asn Glu Ile Val Ile
100 105 110
Val His Asn Phe Ile Lys Asp Lys Tyr Lys Leu Lys Phe Gln Glu Leu
115 120 125
Glu Ser Leu Val His His Pro Ile Asp Tyr Ala Cys Val Val Lys Lys
130 135 140
Ile Gly Asn Glu Thr Asp Leu Ala Leu Val Asp Leu Ala Asp Leu Leu
145 150 155 160
Pro Ser Ala Ile Ile Met Val Val Ser Val Thr Ala Leu Thr Thr Lys
165 170 175
Gly Ser Ala Leu Pro Glu Asp Val Leu Gln Lys Val Leu Glu Ala Cys
180 185 190
Asp Arg Ala Leu Asp Leu Asp Ser Ala Arg Lys Lys Val Leu Glu Phe
195 200 205
Val Glu Ser Lys Met Gly Ser Ile Ala Pro Asn Leu Ser Ala Ile Val
210 215 220
Gly Ser Ala Val Ala Ala Lys Leu Met Gly Thr Ala Gly Gly Leu Ser
225 230 235 240
Ala Leu Ala Lys Met Pro Ala Cys Asn Val Gln Val Leu Gly His Lys
245 250 255
Arg Lys Asn Leu Ala Gly Phe Ser Ser Ala Thr Ser Gln Ser Arg Val
260 265 270
Gly Tyr Leu Glu Gln Thr Glu Ile Tyr Gln Ser Thr Pro Pro Gly Leu
275 280 285
Gln Ala Arg Ala Gly Arg Leu Val Ala Ala Lys Ser Thr Leu Ala Ala
290 295 300
Arg Val Asp Ala Thr Arg Gly Asp Pro Leu Gly Ile Ser Gly Lys Ala
305 310 315 320
Phe Arg Glu Glu Ile Arg Lys Lys Ile Glu Lys Trp Gln Glu Pro Pro
325 330 335
Pro Ala Arg Gln Pro Lys Pro Leu Pro Val Pro Asp Ser Glu Pro Lys
340 345 350
Lys Arg Arg Gly Gly Arg Arg Leu Arg Lys Met Lys Glu Arg Tyr Gln
355 360 365
Val Thr Asp Met Arg Lys Leu Ala Asn Arg Met Ala Phe Gly Thr Pro
370 375 380
Glu Glu Ser Ser Leu Gly Asp Gly Leu Gly Glu Gly Tyr Gly Met Leu
385 390 395 400
Gly Gln Ala Gly Ser Asn Arg Leu Arg Val Ser Ser Val Pro Ser Lys
405 410 415
Leu Lys Ile Asn Ala Lys Val Ala Lys Lys Leu Lys Glu Arg Gln Tyr
420 425 430
Ala Gly Gly Ala Thr Thr Ser Gly Leu Thr Ser Ser Leu Ala Phe Thr
435 440 445
Pro Val Gln Gly Ile Glu Leu Cys Asn Pro Gln Gln Ala Leu Gly Leu
450 455 460
Gly Ser Gly Thr Gln Ser Thr Tyr Phe Ser Glu Ser Gly Thr Phe Ser
465 470 475 480
Lys Leu Lys Lys Ile
485

<210> 5
 <211> 1344
 <212> DNA
 <213> 鼠耳芥

<220>
 <221> CDS
 <222> (1)..(1344)

<400> 5

```

atg gag aac ctt acc cta gtt tct tgc tca gct tct tct cca aag ctg   48
Met Glu Asn Leu Thr Leu Val Ser Cys Ser Ala Ser Ser Pro Lys Leu
   1                               5                               10                               15

tta att gga tgc aat ttc act tcc tcg ctg aaa aac cct act ggg ttt   96
Leu Ile Gly Cys Asn Phe Thr Ser Ser Leu Lys Asn Pro Thr Gly Phe
                               20                               25                               30

tct cgt cgg act cct aat att gtc ctc cgg tgt tcc aaa ata tct gcc  144
Ser Arg Arg Thr Pro Asn Ile Val Leu Arg Cys Ser Lys Ile Ser Ala
                               35                               40                               45

tct gct caa tct caa tct ccc tct tcg cgt ccg gag aac act gga gaa  192
Ser Ala Gln Ser Gln Ser Pro Ser Ser Arg Pro Glu Asn Thr Gly Glu
   50                               55                               60

atc gtg gtt gtg aaa cag aga agc aaa gct ttt gca agt ata ttt tct  240
Ile Val Val Val Lys Gln Arg Ser Lys Ala Phe Ala Ser Ile Phe Ser
   65                               70                               75                               80

tcg agt cgt gat caa cag aca act tct gtt gct tcc cct agt gtg cct  288
Ser Ser Arg Asp Gln Gln Thr Thr Ser Val Ala Ser Pro Ser Val Pro
                               85                               90                               95

gtg cca cca cca tct tca tca acc ata gga tca cca ctt ttc tgg att  336
Val Pro Pro Pro Ser Ser Ser Thr Ile Gly Ser Pro Leu Phe Trp Ile
                               100                              105                              110

ggt gtt ggt gtt ggt cta tca gct ttg ttc tca tat gta act tca aat  384
Gly Val Gly Val Gly Leu Ser Ala Leu Phe Ser Tyr Val Thr Ser Asn
                               115                              120                              125

tta aag aaa tat gca atg caa aca gct atg aag acg atg atg aac caa  432
Leu Lys Lys Tyr Ala Met Gln Thr Ala Met Lys Thr Met Met Asn Gln
                               130                              135                              140

atg aat acg caa aat agc cag ttt aat aat tct gga ttc cca tca gga  480
Met Asn Thr Gln Asn Ser Gln Phe Asn Asn Ser Gly Phe Pro Ser Gly
                               145                              150                              155                              160

tca cct ttt ccg ttt cca ttt cct cct caa aca agt cct gct tcc tcg  528
Ser Pro Phe Pro Phe Pro Phe Pro Pro Gln Thr Ser Pro Ala Ser Ser
                               165                              170                              175

cca ttc caa tct caa tcc cag tct tca ggt gct acc gtt gat gtg aca  576

```

Pro	Phe	Gln	Ser	Gln	Ser	Gln	Ser	Ser	Gly	Ala	Thr	Val	Asp	Val	Thr	
		180						185					190			
gcg	aca	aaa	gta	gag	aca	cct	cct	tca	act	aaa	ccg	aaa	cct	aca	cct	624
Ala	Thr	Lys	Val	Glu	Thr	Pro	Pro	Ser	Thr	Lys	Pro	Lys	Pro	Thr	Pro	
		195					200					205				
gca	aag	gat	ata	gag	gtg	gat	aag	cca	agt	gtt	gtc	tta	gag	gca	agc	672
Ala	Lys	Asp	Ile	Glu	Val	Asp	Lys	Pro	Ser	Val	Val	Leu	Glu	Ala	Ser	
	210					215				220						
aaa	gag	aag	aaa	gaa	gaa	aag	aac	tat	gcc	ttt	gaa	gac	att	tca	ccc	720
Lys	Glu	Lys	Lys	Glu	Glu	Lys	Asn	Tyr	Ala	Phe	Glu	Asp	Ile	Ser	Pro	
225				230					235						240	
gag	gaa	acc	aca	aaa	gaa	agc	cca	ttt	agc	aac	tat	gca	gaa	gtc	tct	768
Glu	Glu	Thr	Thr	Lys	Glu	Ser	Pro	Phe	Ser	Asn	Tyr	Ala	Glu	Val	Ser	
			245					250						255		
gaa	act	aat	tcc	ccc	aaa	gaa	act	cgc	ttg	ttt	gag	gat	gtc	ttg	caa	816
Glu	Thr	Asn	Ser	Pro	Lys	Glu	Thr	Arg	Leu	Phe	Glu	Asp	Val	Leu	Gln	
		260						265					270			
aat	gga	gct	ggt	ccg	gca	aat	ggt	gcc	act	gct	tca	gag	gtt	ttt	caa	864
Asn	Gly	Ala	Gly	Pro	Ala	Asn	Gly	Ala	Thr	Ala	Ser	Glu	Val	Phe	Gln	
		275					280					285				
tct	ttg	ggt	ggt	ggg	aaa	gga	ggg	ccg	ggt	tta	tct	gta	gaa	gct	tta	912
Ser	Leu	Gly	Gly	Gly	Lys	Gly	Gly	Pro	Gly	Leu	Ser	Val	Glu	Ala	Leu	
	290					295					300					
gag	aaa	atg	atg	gaa	gat	cca	aca	gtc	cag	aag	atg	gtt	tac	cca	tac	960
Glu	Lys	Met	Met	Glu	Asp	Pro	Thr	Val	Gln	Lys	Met	Val	Tyr	Pro	Tyr	
305				310					315						320	
ttg	cct	gag	gag	atg	agg	aac	cca	gaa	act	ttc	aaa	tgg	atg	ctt	aaa	1008
Leu	Pro	Glu	Glu	Met	Arg	Asn	Pro	Glu	Thr	Phe	Lys	Trp	Met	Leu	Lys	
			325					330						335		
aat	cct	cag	tac	cgt	caa	caa	cta	cag	gac	atg	ttg	aat	aat	atg	agt	1056
Asn	Pro	Gln	Tyr	Arg	Gln	Gln	Leu	Gln	Asp	Met	Leu	Asn	Asn	Met	Ser	
		340					345					350				
ggg	agt	ggt	gaa	tgg	gac	aag	cga	atg	aca	gat	aca	ttg	aag	aat	ttt	1104
Gly	Ser	Gly	Glu	Trp	Asp	Lys	Arg	Met	Thr	Asp	Thr	Leu	Lys	Asn	Phe	
		355				360						365				
gac	ctg	aat	agt	cct	gaa	gtg	aag	caa	caa	ttc	aat	caa	ata	gga	cta	1152
Asp	Leu	Asn	Ser	Pro	Glu	Val	Lys	Gln	Gln	Phe	Asn	Gln	Ile	Gly	Leu	
	370					375				380						
act	cca	gaa	gaa	gtc	ata	tct	aag	atc	atg	gag	aac	cct	gat	gtt	gcc	1200
Thr	Pro	Glu	Glu	Val	Ile	Ser	Lys	Ile	Met	Glu	Asn	Pro	Asp	Val	Ala	
385				390					395						400	

atg gca ttc cag aat cct aga gtc caa gca gcg tta atg gaa tgc tca 1248
 Met Ala Phe Gln Asn Pro Arg Val Gln Ala Ala Leu Met Glu Cys Ser
 405 410 415

gag aac cca atg aac atc atg aag tac caa aac gac aaa gag gta atg 1296
 Glu Asn Pro Met Asn Ile Met Lys Tyr Gln Asn Asp Lys Glu Val Met
 420 425 430

gat gtg ttc aac aag ata tcg cag ctc ttc cca gga atg acg ggt tga 1344
 Asp Val Phe Asn Lys Ile Ser Gln Leu Phe Pro Gly Met Thr Gly
 435 440 445

<210> 6

<211> 447

<212> PRT

<213> 鼠耳芥

<400> 6

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

Ser Leu Gly Gly Gly Lys Gly Gly Pro Gly Leu Ser Val Glu Ala Leu
 290 295 300
 Glu Lys Met Met Glu Asp Pro Thr Val Gln Lys Met Val Tyr Pro Tyr
 305 310 315 320
 Leu Pro Glu Glu Met Arg Asn Pro Glu Thr Phe Lys Trp Met Leu Lys
 325 330 335
 Asn Pro Gln Tyr Arg Gln Gln Leu Gln Asp Met Leu Asn Asn Met Ser
 340 345 350
 Gly Ser Gly Glu Trp Asp Lys Arg Met Thr Asp Thr Leu Lys Asn Phe
 355 360 365
 Asp Leu Asn Ser Pro Glu Val Lys Gln Gln Phe Asn Gln Ile Gly Leu
 370 375 380
 Thr Pro Glu Glu Val Ile Ser Lys Ile Met Glu Asn Pro Asp Val Ala
 385 390 395 400
 Met Ala Phe Gln Asn Pro Arg Val Gln Ala Ala Leu Met Glu Cys Ser
 405 410 415
 Glu Asn Pro Met Asn Ile Met Lys Tyr Gln Asn Asp Lys Glu Val Met
 420 425 430
 Asp Val Phe Asn Lys Ile Ser Gln Leu Phe Pro Gly Met Thr Gly
 435 440 445

<210> 7
 <211> 2163
 <212> DNA
 <213> 鼠耳芥

<220>
 <221> CDS
 <222> (1)..(2163)

<400> 7
 atg tct agg gag gat ttt agt gat aca ctt cga gta ctt gtt gca act 48
 Met Ser Arg Glu Asp Phe Ser Asp Thr Leu Arg Val Leu Val Ala Thr
 1 5 10 15
 gat tgc cac ttg ggc tac atg gag aag gat gaa att agg cgg cat gat 96
 Asp Cys His Leu Gly Tyr Met Glu Lys Asp Glu Ile Arg Arg His Asp
 20 25 30
 tca ttt aag gct ttc gaa gag ata tgt tct ata gct gag gag aaa cag 144
 Ser Phe Lys Ala Phe Glu Glu Ile Cys Ser Ile Ala Glu Glu Lys Gln
 35 40 45
 gtg gac ttc tta ctc ctc gga ggt gat ctt ttt cat gag aat aaa ccc 192
 Val Asp Phe Leu Leu Leu Gly Gly Asp Leu Phe His Glu Asn Lys Pro
 50 55 60
 tct aga act acg tta gtt aaa gcc att gaa att ctt cgt cgc cac tgt 240
 Ser Arg Thr Thr Leu Val Lys Ala Ile Glu Ile Leu Arg Arg His Cys
 65 70 75 80
 ctg aat gat aaa cca gtg cag ttt caa gta gtc agc gac cag aca gta 288
 Leu Asn Asp Lys Pro Val Gln Phe Gln Val Val Ser Asp Gln Thr Val

	85		90		95		
aat ttt cag aat gcg ttt ggt caa gtc aat tac gag gat cca cac ttc						336	
Asn Phe Gln Asn Ala Phe Gly Gln Val Asn Tyr Glu Asp Pro His Phe	100		105		110		
aat gta ggc ttg ccc gtg ttc agt att cat gga aac cat gat gat cca						384	
Asn Val Gly Leu Pro Val Phe Ser Ile His Gly Asn His Asp Asp Pro	115		120		125		
gcc gga gtg gac aat ctt tct gca att gat att ctt tcc gca tgc aac						432	
Ala Gly Val Asp Asn Leu Ser Ala Ile Asp Ile Leu Ser Ala Cys Asn	130		135		140		
ctt gtg aac tat ttt gga aag atg gtt ctt ggt ggt tct ggt gtt ggc						480	
Leu Val Asn Tyr Phe Gly Lys Met Val Leu Gly Gly Ser Gly Val Gly	145	150		155	160		
cag att act ctc tac cct ata ctt atg aag aag ggc tca aca acc gtg						528	
Gln Ile Thr Leu Tyr Pro Ile Leu Met Lys Lys Gly Ser Thr Thr Val	165		170		175		
gct ctc tat ggt tta gga aac atc agg gat gaa cgt ctc aat aga atg						576	
Ala Leu Tyr Gly Leu Gly Asn Ile Arg Asp Glu Arg Leu Asn Arg Met	180		185		190		
ttt cag acc cca cat gct gtc caa tgg atg agg cct gaa gtt caa gaa						624	
Phe Gln Thr Pro His Ala Val Gln Trp Met Arg Pro Glu Val Gln Glu	195		200		205		
gga tgt gat gtt tct gac tgg ttc aac att ctg gtg ctt cat caa aat						672	
Gly Cys Asp Val Ser Asp Trp Phe Asn Ile Leu Val Leu His Gln Asn	210		215		220		
agg gtg aaa tca aac ccc aaa aat gca ata agt gag cac ttt ctt cca						720	
Arg Val Lys Ser Asn Pro Lys Asn Ala Ile Ser Glu His Phe Leu Pro	225	230		235	240		
cgt ttc ctc gac ttc att gtg tgg ggc cat gag cat gaa tgc cta atc						768	
Arg Phe Leu Asp Phe Ile Val Trp Gly His Glu His Glu Cys Leu Ile	245		250		255		
gac ccc cag gag gta tct gga atg ggc ttc cac atc aca caa cca gga						816	
Asp Pro Gln Glu Val Ser Gly Met Gly Phe His Ile Thr Gln Pro Gly	260		265		270		
tct tct gtg gca aca tca ctt att gat ggg gaa tcg aag cca aaa cat						864	
Ser Ser Val Ala Thr Ser Leu Ile Asp Gly Glu Ser Lys Pro Lys His	275		280		285		
gtt ctt ctc tta gaa atc aag gga aat caa tat cgt cct acg aag ata						912	
Val Leu Leu Leu Glu Ile Lys Gly Asn Gln Tyr Arg Pro Thr Lys Ile	290	295		300			
cct ttg aca tct gtg agg cct ttt gag tat aca gag att gtt tta aag						960	

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

gat gaa agt gat att gat ccc aat gat caa aac tca att ctg gaa cac 1008
 Asp Glu Ser Asp Ile Asp Pro Asn Asp Gln Asn Ser Ile Leu Glu His
 325 330 335

ttg gat aaa gtg gtc aga aat cta ata gag aaa gct agc aaa aaa gct 1056
 Leu Asp Lys Val Val Arg Asn Leu Ile Glu Lys Ala Ser Lys Lys Ala
 340 345 350

gtt aac aga tca gag atc aaa ctc cca ttg gtt cga atc aag gta gat 1104
 Val Asn Arg Ser Glu Ile Lys Leu Pro Leu Val Arg Ile Lys Val Asp
 355 360 365

tat tct gga ttt atg acg ata aat cct caa aga ttt gga cag aaa tat 1152
 Tyr Ser Gly Phe Met Thr Ile Asn Pro Gln Arg Phe Gly Gln Lys Tyr
 370 375 380

gtg gga aag gtt gca aat ccc cag gac att ttg ata ttt tcc aag gct 1200
 Val Gly Lys Val Ala Asn Pro Gln Asp Ile Leu Ile Phe Ser Lys Ala
 385 390 395 400

tct aag aag ggt cgg agc gaa gcc aac atc gat gat tct gag cgg ctt 1248
 Ser Lys Lys Gly Arg Ser Glu Ala Asn Ile Asp Asp Ser Glu Arg Leu
 405 410 415

cgt cca gaa gaa ctg aac cag cag aat ata gaa gct tta gta gct gaa 1296
 Arg Pro Glu Glu Leu Asn Gln Gln Asn Ile Glu Ala Leu Val Ala Glu
 420 425 430

agc aac ctg aaa atg gag atc ctt cca gtt aac gat ctg gat gtt gct 1344
 Ser Asn Leu Lys Met Glu Ile Leu Pro Val Asn Asp Leu Asp Val Ala
 435 440 445

ctt cac aat ttt gtg aac aag gat gat aaa cta gcc ttc tac tca tgc 1392
 Leu His Asn Phe Val Asn Lys Asp Asp Lys Leu Ala Phe Tyr Ser Cys
 450 455 460

gtt cag tac aat ctt caa gag act cgt ggt aaa ctt gca aag gat tca 1440
 Val Gln Tyr Asn Leu Gln Glu Thr Arg Gly Lys Leu Ala Lys Asp Ser
 465 470 475 480

gat gcc aag aaa ttt gag gaa gat gac ttg att ctt aaa gtg gga gag 1488
 Asp Ala Lys Lys Phe Glu Glu Asp Asp Leu Ile Leu Lys Val Gly Glu
 485 490 495

tgc tta gag gaa cgc ttg aaa gat agg tcc act cga ccc act ggt tcc 1536
 Cys Leu Glu Glu Arg Leu Lys Asp Arg Ser Thr Arg Pro Thr Gly Ser
 500 505 510

tca cag ttt tta tcc act gga ttg act tca gag aat ttg aca aaa gga 1584
 Ser Gln Phe Leu Ser Thr Gly Leu Thr Ser Glu Asn Leu Thr Lys Gly
 515 520 525

```

agc agt ggc atc gcg aat gct tcg ttc agt gat gat gaa gac aca act 1632
Ser Ser Gly Ile Ala Asn Ala Ser Phe Ser Asp Asp Glu Asp Thr Thr
      530                      535                      540

cag atg tct ggt tta gct cct ccc act aga gga cga aga ggt tca tcc 1680
Gln Met Ser Gly Leu Ala Pro Pro Thr Arg Gly Arg Arg Gly Ser Ser
545                      550                      555                      560

act gct aat aca act cgt ggt aga gct aaa gcc cca acc aga gga cga 1728
Thr Ala Asn Thr Thr Arg Gly Arg Ala Lys Ala Pro Thr Arg Gly Arg
      565                      570                      575

ggc cgt ggt aag gcc tca agt gcg atg aag caa acc act ctt gat agt 1776
Gly Arg Gly Lys Ala Ser Ser Ala Met Lys Gln Thr Thr Leu Asp Ser
      580                      585                      590

tct ctt ggt ttc cgc cag tct caa aga tct gct tcg gct gct gct tca 1824
Ser Leu Gly Phe Arg Gln Ser Gln Arg Ser Ala Ser Ala Ala Ser
      595                      600                      605

gct gcc ttc aaa agt gct tcc acc att gga gaa gat gat gta gat tct 1872
Ala Ala Phe Lys Ser Ala Ser Thr Ile Gly Glu Asp Asp Val Asp Ser
      610                      615                      620

cct tca agc gaa gaa gtc gag cct gaa gat ttt aac aaa cct gac agc 1920
Pro Ser Ser Glu Glu Val Glu Pro Glu Asp Phe Asn Lys Pro Asp Ser
625                      630                      635                      640

agt tcg gag gac gat gag agc act aaa ggc aaa gga cgt aaa aga cca 1968
Ser Ser Glu Asp Asp Glu Ser Thr Lys Gly Lys Gly Arg Lys Arg Pro
      645                      650                      655

gct act act aag aga ggc aga ggt aga ggt tct ggg act tca aaa cgt 2016
Ala Thr Thr Lys Arg Gly Arg Gly Arg Gly Ser Gly Thr Ser Lys Arg
      660                      665                      670

ggt aga aaa aac gaa agc tct tct tca ctt aat agg cta ctc agt agc 2064
Gly Arg Lys Asn Glu Ser Ser Ser Ser Leu Asn Arg Leu Leu Ser Ser
      675                      680                      685

aaa gac gat gac gag gac gaa gat gat gaa gac aga gaa aag aag ctt 2112
Lys Asp Asp Asp Glu Asp Glu Asp Asp Glu Asp Arg Glu Lys Lys Leu
      690                      695                      700

aac aaa tct cag cct cgg gtt aca agg aac tat gga gct cta aga aga 2160
Asn Lys Ser Gln Pro Arg Val Thr Arg Asn Tyr Gly Ala Leu Arg Arg
705                      710                      715                      720

taa 2163

```

```

<210> 8
<211> 720
<212> PRT

```

<213> 鼠耳芥

<400> 8

Met Ser Arg Glu Asp Phe Ser Asp Thr Leu Arg Val Leu Val Ala Thr
 1 5 10 15
 Asp Cys His Leu Gly Tyr Met Glu Lys Asp Glu Ile Arg Arg His Asp
 20 25 30
 Ser Phe Lys Ala Phe Glu Glu Ile Cys Ser Ile Ala Glu Glu Lys Gln
 35 40 45
 Val Asp Phe Leu Leu Leu Gly Asp Leu Phe His Glu Asn Lys Pro
 50 55 60
 Ser Arg Thr Thr Leu Val Lys Ala Ile Glu Ile Leu Arg Arg His Cys
 65 70 75 80
 Leu Asn Asp Lys Pro Val Gln Phe Gln Val Val Ser Asp Gln Thr Val
 85 90 95
 Asn Phe Gln Asn Ala Phe Gly Gln Val Asn Tyr Glu Asp Pro His Phe
 100 105 110
 Asn Val Gly Leu Pro Val Phe Ser Ile His Gly Asn His Asp Asp Pro
 115 120 125
 Ala Gly Val Asp Asn Leu Ser Ala Ile Asp Ile Leu Ser Ala Cys Asn
 130 135 140
 Leu Val Asn Tyr Phe Gly Lys Met Val Leu Gly Gly Ser Gly Val Gly
 145 150 155 160
 Gln Ile Thr Leu Tyr Pro Ile Leu Met Lys Lys Gly Ser Thr Thr Val
 165 170 175
 Ala Leu Tyr Gly Leu Gly Asn Ile Arg Asp Glu Arg Leu Asn Arg Met
 180 185 190
 Phe Gln Thr Pro His Ala Val Gln Trp Met Arg Pro Glu Val Gln Glu
 195 200 205
 Gly Cys Asp Val Ser Asp Trp Phe Asn Ile Leu Val Leu His Gln Asn
 210 215 220
 Arg Val Lys Ser Asn Pro Lys Asn Ala Ile Ser Glu His Phe Leu Pro
 225 230 235 240
 Arg Phe Leu Asp Phe Ile Val Trp Gly His Glu His Glu Cys Leu Ile
 245 250 255
 Asp Pro Gln Glu Val Ser Gly Met Gly Phe His Ile Thr Gln Pro Gly
 260 265 270
 Ser Ser Val Ala Thr Ser Leu Ile Asp Gly Glu Ser Lys Pro Lys His
 275 280 285
 Val Leu Leu Leu Glu Ile Lys Gly Asn Gln Tyr Arg Pro Thr Lys Ile
 290 295 300
 Pro Leu Thr Ser Val Arg Pro Phe Glu Tyr Thr Glu Ile Val Leu Lys
 305 310 315 320
 Asp Glu Ser Asp Ile Asp Pro Asn Asp Gln Asn Ser Ile Leu Glu His
 325 330 335
 Leu Asp Lys Val Val Arg Asn Leu Ile Glu Lys Ala Ser Lys Lys Ala
 340 345 350
 Val Asn Arg Ser Glu Ile Lys Leu Pro Leu Val Arg Ile Lys Val Asp
 355 360 365
 Tyr Ser Gly Phe Met Thr Ile Asn Pro Gln Arg Phe Gly Gln Lys Tyr
 370 375 380
 Val Gly Lys Val Ala Asn Pro Gln Asp Ile Leu Ile Phe Ser Lys Ala
 385 390 395 400
 Ser Lys Lys Gly Arg Ser Glu Ala Asn Ile Asp Asp Ser Glu Arg Leu
 405 410 415

Arg Pro Glu Glu Leu Asn Gln Gln Asn Ile Glu Ala Leu Val Ala Glu
 420 425 430
 Ser Asn Leu Lys Met Glu Ile Leu Pro Val Asn Asp Leu Asp Val Ala
 435 440 445
 Leu His Asn Phe Val Asn Lys Asp Asp Lys Leu Ala Phe Tyr Ser Cys
 450 455 460
 Val Gln Tyr Asn Leu Gln Glu Thr Arg Gly Lys Leu Ala Lys Asp Ser
 465 470 475 480
 Asp Ala Lys Lys Phe Glu Glu Asp Asp Leu Ile Leu Lys Val Gly Glu
 485 490 495
 Cys Leu Glu Glu Arg Leu Lys Asp Arg Ser Thr Arg Pro Thr Gly Ser
 500 505 510
 Ser Gln Phe Leu Ser Thr Gly Leu Thr Ser Glu Asn Leu Thr Lys Gly
 515 520 525
 Ser Ser Gly Ile Ala Asn Ala Ser Phe Ser Asp Asp Glu Asp Thr Thr
 530 535 540
 Gln Met Ser Gly Leu Ala Pro Pro Thr Arg Gly Arg Arg Gly Ser Ser
 545 550 555 560
 Thr Ala Asn Thr Thr Arg Gly Arg Ala Lys Ala Pro Thr Arg Gly Arg
 565 570 575
 Gly Arg Gly Lys Ala Ser Ser Ala Met Lys Gln Thr Thr Leu Asp Ser
 580 585 590
 Ser Leu Gly Phe Arg Gln Ser Gln Arg Ser Ala Ser Ala Ala Ala Ser
 595 600 605
 Ala Ala Phe Lys Ser Ala Ser Thr Ile Gly Glu Asp Asp Val Asp Ser
 610 615 620
 Pro Ser Ser Glu Glu Val Glu Pro Glu Asp Phe Asn Lys Pro Asp Ser
 625 630 635 640
 Ser Ser Glu Asp Asp Glu Ser Thr Lys Gly Lys Gly Arg Lys Arg Pro
 645 650 655
 Ala Thr Thr Lys Arg Gly Arg Gly Arg Gly Ser Gly Thr Ser Lys Arg
 660 665 670
 Gly Arg Lys Asn Glu Ser Ser Ser Ser Leu Asn Arg Leu Leu Ser Ser
 675 680 685
 Lys Asp Asp Asp Glu Asp Glu Asp Asp Glu Asp Arg Glu Lys Lys Leu
 690 695 700
 Asn Lys Ser Gln Pro Arg Val Thr Arg Asn Tyr Gly Ala Leu Arg Arg
 705 710 715 720

<210> 9

<211> 1434

<212> DNA

<213> 鼠耳芥

<220>

<221> CDS

<222> (1)..(1434)

<400> 9

atg atg aca tta aac tca cta tct cca gct gaa tcc aaa gct att tct 48
 Met Met Thr Leu Asn Ser Leu Ser Pro Ala Glu Ser Lys Ala Ile Ser
 1 5 10 15

ttc ttg gat acc tcc agg ttc aat cca atc cct aaa ctc tca ggt ggg 96
 Phe Leu Asp Thr Ser Arg Phe Asn Pro Ile Pro Lys Leu Ser Gly Gly
 20 25 30

ttt agt ttg agg agg agg gat caa ggg aga ggt ttt gga aaa ggt gtt 144
 Phe Ser Leu Arg Arg Arg Asp Gln Gly Arg Gly Phe Gly Lys Gly Val
 35 40 45

aag tgt tca gtg aaa gtg cag cag caa caa caa cct cct cca gca tgg 192
 Lys Cys Ser Val Lys Val Gln Gln Gln Gln Gln Pro Pro Pro Ala Trp
 50 55 60

cct ggg aga gct gtt cct gag gcg cct cgt caa tct tgg gat gga cca 240
 Pro Gly Arg Ala Val Pro Glu Ala Pro Arg Gln Ser Trp Asp Gly Pro
 65 70 75 80

aaa ccc atc tct atc gtt gga tct act ggt tcc atc ggc act cag aca 288
 Lys Pro Ile Ser Ile Val Gly Ser Thr Gly Ser Ile Gly Thr Gln Thr
 85 90 95

ttg gat att gtg gct gag aat cct gac aaa ttt aga gtt gtg gct cta 336
 Leu Asp Ile Val Ala Glu Asn Pro Asp Lys Phe Arg Val Val Ala Leu
 100 105 110

gct gct ggt tcg aat gtt act cta ctt gct gat cag gta agg aga ttt 384
 Ala Ala Gly Ser Asn Val Thr Leu Leu Ala Asp Gln Val Arg Arg Phe
 115 120 125

aag cct gcg ttg gtt gct gtt aga aac gag tca ctg att aat gag ctt 432
 Lys Pro Ala Leu Val Ala Val Arg Asn Glu Ser Leu Ile Asn Glu Leu
 130 135 140

aaa gag gct tta gct gat ttg gac tat aaa ccc gag att att cca gga 480
 Lys Glu Ala Leu Ala Asp Leu Asp Tyr Lys Pro Glu Ile Ile Pro Gly
 145 150 155 160

gag cta gga gtg att gag gtt gcc cga cat cct gaa gct gta acc gtt 528
 Glu Leu Gly Val Ile Glu Val Ala Arg His Pro Glu Ala Val Thr Val
 165 170 175

gtt acc gga ata gta ggt tgt gcg gga ctg aag cct acg gtt gct gca 576
 Val Thr Gly Ile Val Gly Cys Ala Gly Leu Lys Pro Thr Val Ala Ala
 180 185 190

att gaa gca gga aag gac att gct ctt gca aac aaa gag aca tta atc 624
 Ile Glu Ala Gly Lys Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile
 195 200 205

gca ggt ggt cct ttc gtg ctt ccg ctt gcc aac aaa cat aat gta aag 672
 Ala Gly Gly Pro Phe Val Leu Pro Leu Ala Asn Lys His Asn Val Lys
 210 215 220

att ctt ccg gca gat tca gaa cat tct gcc ata ttt cag tgt att caa 720
 Ile Leu Pro Ala Asp Ser Glu His Ser Ala Ile Phe Gln Cys Ile Gln
 225 230 235 240

ggt ttg cct gaa ggc gct ctg cgc aag ata atc ttg act gca tct ggt 768
 Gly Leu Pro Glu Gly Ala Leu Arg Lys Ile Ile Leu Thr Ala Ser Gly
 245 250 255

gga gct ttt agg gat tgg cct gtc gaa aag cta aag gaa gtt aaa gta 816
 Gly Ala Phe Arg Asp Trp Pro Val Glu Lys Leu Lys Glu Val Lys Val
 260 265 270

gcg gat gcg ttg aag cat cca aac tgg aac atg gga aag aaa atc act 864
 Ala Asp Ala Leu Lys His Pro Asn Trp Asn Met Gly Lys Lys Ile Thr
 275 280 285

gtg gac tct gct acg ctt ttc aac aag ggt ctt gag gtc att gaa gcg 912
 Val Asp Ser Ala Thr Leu Phe Asn Lys Gly Leu Glu Val Ile Glu Ala
 290 295 300

cat tat ttg ttt gga gct gag tat gac gat ata gag att gtc att cat 960
 His Tyr Leu Phe Gly Ala Glu Tyr Asp Asp Ile Glu Ile Val Ile His
 305 310 315 320

cct caa agt atc ata cat tcc atg att gaa aca cag gat tca tct gtg 1008
 Pro Gln Ser Ile Ile His Ser Met Ile Glu Thr Gln Asp Ser Ser Val
 325 330 335

ctt gct caa ttg ggt tgg cct gat atg cgt tta ccg att ctc tac acc 1056
 Leu Ala Gln Leu Gly Trp Pro Asp Met Arg Leu Pro Ile Leu Tyr Thr
 340 345 350

atg tca tgg ccc gat aga gtt cct tgt tct gaa gta act tgg cct aga 1104
 Met Ser Trp Pro Asp Arg Val Pro Cys Ser Glu Val Thr Trp Pro Arg
 355 360 365

ctt gac ctt tgc aaa ctc ggt tca ttg act ttc aag aaa cca gac aat 1152
 Leu Asp Leu Cys Lys Leu Gly Ser Leu Thr Phe Lys Lys Pro Asp Asn
 370 375 380

gtg aaa tac cca tcc atg gat ctt gct tat gct gct gga cga gct gga 1200
 Val Lys Tyr Pro Ser Met Asp Leu Ala Tyr Ala Ala Gly Arg Ala Gly
 385 390 395 400

ggc aca atg act gga gtt ctc agc gcc gcc aat gag aaa gct gtt gaa 1248
 Gly Thr Met Thr Gly Val Leu Ser Ala Ala Asn Glu Lys Ala Val Glu
 405 410 415

atg ttt att gat gaa aag ata agc tat ttg gat atc ttc aag gtt gtg 1296
 Met Phe Ile Asp Glu Lys Ile Ser Tyr Leu Asp Ile Phe Lys Val Val
 420 425 430

gaa tta aca tgc gat aaa cat cga aac gag ttg gta aca tca ccg tct 1344
 Glu Leu Thr Cys Asp Lys His Arg Asn Glu Leu Val Thr Ser Pro Ser
 435 440 445

ctt gaa gag att gtt cac tat gac ttg tgg gca cgt gaa tat gcc gcg 1392
 Leu Glu Glu Ile Val His Tyr Asp Leu Trp Ala Arg Glu Tyr Ala Ala

450

455

460

gat gtg cag ctt tct tct ggt gct agg cca gtt cat gca tga
 Asp Val Gln Leu Ser Ser Gly Ala Arg Pro Val His Ala
 465 470 475

1434

<210> 10
 <211> 477
 <212> PRT
 <213> 鼠耳芥

<400> 10

Met Met Thr Leu Asn Ser Leu Ser Pro Ala Glu Ser Lys Ala Ile Ser
 1 5 10 15
 Phe Leu Asp Thr Ser Arg Phe Asn Pro Ile Pro Lys Leu Ser Gly Gly
 20 25 30
 Phe Ser Leu Arg Arg Arg Asp Gln Gly Arg Gly Phe Gly Lys Gly Val
 35 40 45
 Lys Cys Ser Val Lys Val Gln Gln Gln Gln Gln Pro Pro Pro Ala Trp
 50 55 60
 Pro Gly Arg Ala Val Pro Glu Ala Pro Arg Gln Ser Trp Asp Gly Pro
 65 70 75 80
 Lys Pro Ile Ser Ile Val Gly Ser Thr Gly Ser Ile Gly Thr Gln Thr
 85 90 95
 Leu Asp Ile Val Ala Glu Asn Pro Asp Lys Phe Arg Val Val Ala Leu
 100 105 110
 Ala Ala Gly Ser Asn Val Thr Leu Leu Ala Asp Gln Val Arg Arg Phe
 115 120 125
 Lys Pro Ala Leu Val Ala Val Arg Asn Glu Ser Leu Ile Asn Glu Leu
 130 135 140
 Lys Glu Ala Leu Ala Asp Leu Asp Tyr Lys Pro Glu Ile Ile Pro Gly
 145 150 155 160
 Glu Leu Gly Val Ile Glu Val Ala Arg His Pro Glu Ala Val Thr Val
 165 170 175
 Val Thr Gly Ile Val Gly Cys Ala Gly Leu Lys Pro Thr Val Ala Ala
 180 185 190
 Ile Glu Ala Gly Lys Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile
 195 200 205
 Ala Gly Gly Pro Phe Val Leu Pro Leu Ala Asn Lys His Asn Val Lys
 210 215 220
 Ile Leu Pro Ala Asp Ser Glu His Ser Ala Ile Phe Gln Cys Ile Gln
 225 230 235 240
 Gly Leu Pro Glu Gly Ala Leu Arg Lys Ile Ile Leu Thr Ala Ser Gly
 245 250 255
 Gly Ala Phe Arg Asp Trp Pro Val Glu Lys Leu Lys Glu Val Lys Val
 260 265 270
 Ala Asp Ala Leu Lys His Pro Asn Trp Asn Met Gly Lys Lys Ile Thr
 275 280 285
 Val Asp Ser Ala Thr Leu Phe Asn Lys Gly Leu Glu Val Ile Glu Ala
 290 295 300
 His Tyr Leu Phe Gly Ala Glu Tyr Asp Asp Ile Glu Ile Val Ile His
 305 310 315 320
 Pro Gln Ser Ile Ile His Ser Met Ile Glu Thr Gln Asp Ser Ser Val
 325 330 335

```

Leu Ala Gln Leu Gly Trp Pro Asp Met Arg Leu Pro Ile Leu Tyr Thr
          340          345          350
Met Ser Trp Pro Asp Arg Val Pro Cys Ser Glu Val Thr Trp Pro Arg
          355          360          365
Leu Asp Leu Cys Lys Leu Gly Ser Leu Thr Phe Lys Lys Pro Asp Asn
          370          375          380
Val Lys Tyr Pro Ser Met Asp Leu Ala Tyr Ala Ala Gly Arg Ala Gly
          385          390          395          400
Gly Thr Met Thr Gly Val Leu Ser Ala Ala Asn Glu Lys Ala Val Glu
          405          410          415
Met Phe Ile Asp Glu Lys Ile Ser Tyr Leu Asp Ile Phe Lys Val Val
          420          425          430
Glu Leu Thr Cys Asp Lys His Arg Asn Glu Leu Val Thr Ser Pro Ser
          435          440          445
Leu Glu Glu Ile Val His Tyr Asp Leu Trp Ala Arg Glu Tyr Ala Ala
          450          455          460
Asp Val Gln Leu Ser Ser Gly Ala Arg Pro Val His Ala
          465          470          475

```

<210> 11
 <211> 20
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列的描述:
 寡核苷酸

<400> 11
 gcggacatct acatTTTTga

20

<210> 12
 <211> 1353
 <212> DNA
 <213> 鼠耳芥

```

<400> 12
gctgggtaag tagatcgttg catcactatg agatgtctaa gcttcttagg gatcaatatg 60
acgctgaagg cgcttgtatg attatcaaat ctggatctcc aggcgcaaaa tctcagggtca 120
gtttcatcat tctcaaggca cttacagttt ccaactcttt gcttgtaact tagtttctgt 180
ttgttcttaa acatattttg aggatttgca gatatggaca gagcaagttg taagtatgta 240
tatcaaatgg gcagaaaggc taggccaana cgcgcggtg gctgagaaat gtagtttatt 300
gagtaataaa agtggcgtaa gttcagccac gatagagttt gaattcgagt ttgcttatgg 360
ttatctctta ggtgagcgag gtgtgcaccg ccttatcata agttccactt ctaatgagg 420
atacattata agttataact ctctttctcg taactaatca ctttctgtgc cattatcatg 480
gcccgggaaa gaattaaaag aggttttctt tgcgccagga atgttcagcg actgttgata 540
tcataccact attcttgaga gcatctcctg attttgaagt aaaggaagg 600
tatcgtatcc tgcaaaagag gatcacaana tagctgagaa tatggtttgt atccaccata 660
ttccgagtgg agtaacacta caatcttcag gtattcttga gtgtgttgtt agttgttaca 720
ctttggttta ctgcatttta tgcagattat ataacatgag gtttttgatg caggagaaag 780
aaaccggttt gcaaacagga tcaaagctct aaaccggtg aaggcgaagc tacttgtgat 840
agcaaaagag caaaaggttt cggatgtaaa taaaatcgac agcaagaaca ttttgaacc 900

```

```

gcgggaagaa accaggagtt atgtctctaa gggtcacaag atgggtggtg atagaaaaac 960
cggtttagag attctggacc tgaaatcggg cttggatgga aacattggac cactccttgg 1020
agctcatatt agcatgagaa gatcaattga tgcgatttag gcttaatcaa ttggtacttt 1080
aattgctttt tgttttgtat ccaaaaagca acaaatgggt gcttgtgtgt gtatatatat 1140
aaccttcttg tccagaacca tatatgattc taaccatcaa acaagataa gaattggtga 1200
ctatgtgcta tactctacaa tatcaccatg aatacttcaa actagacttt tgataaattt 1260
tgaaacgggt attaccaata aaacgaaaac catgaaactc ttgttttaat tatcagattc 1320
gagaaagtgt tgtacaaaca tagctgagaa ggg                                     1353

```

<210> 13
<211> 184
<212> DNA
<213> 鼠耳芥

```

<400> 13
gcttaatcaa ttggtacttt aattgctttt tggtttgtat ccaaaaagca acaaatggkt 60
gcttgtgtgt gtatatatat aaccttcttg gccagaacca tatatgawtc taaccattaa 120
accaagatta gaattggtga ctaaaaaaaaa agaaaaaaaaa aaaaaaaaaaa aaaaaaaaaa 180
aaaa                                             184

```

<210> 14
<211> 2170
<212> DNA
<213> 鼠耳芥

```

<400> 14
atggtaagcg tttctttaac tctattttct tcattgtttc agttattggc gattgtattc 60
tctgtttatt gtaatcgtat tgtgttaatt ttgatttgac tcactctctc taaagttcaa 120
tttcaaaatt agggattccg agatcataga tattgctttg tttccgagat ttgagttatt 180
cttaagcttg ttttactaac tttcaatatg ttggatttgt tataggcaac tcttgaagat 240
tctttccttg ctgatttggg cgagttatct gacaatgaag cagaattggt gagtgtaaa 300
acacttttga ttactattat ctgtttactt ggaggagcta tgattgtaat ttagtattgt 360
ttgattatac atatgcagga cgagaatgat ggtgatgttg gaaaggaaga agaagatggt 420
gatatggata tggctgattt agagacactt aactatgatg atctcgataa tgttttctaa 480
ctgcagaaga gtcagagata tgctgatatt atgcataaag tagaggaggc tcttgggaaa 540
gattctgatg gagctgagaa aggaactgtc ttggaagatg atcctgagta taagcttatt 600
gtggattgta atcagctttc ggtcgatatt gagaatgaaa tcgttattgt ccacaacttt 660
atcaaaagca agtacaagct taagtttcaa gagcttgagt cgttggttca tcaccctatt 720
gactatgcat gtgttgtgaa gaagattggg aatgagacgg atttggctct tgttgatctc 780
gctgaccttc ttccttcagc tattatcatg gttgtttcag ttactgcttt aactacgaaa 840
gggagtgcac tgccagagga tgttttgcaa aagggtttag aggcttgtga tcgggcttta 900
gatcttgatt ccgcaaggaa gaaggtcctt gagtttgttg aaagtaagat gggatctatt 960
gcacctaate tttctgctat tgttgggagt gctgttcag ccaaactcat ggggactgct 1020
ggaggtttgt cagcacttgc taaaatgcct gcgtgtaatg ttcaagttct tggccacaag 1080
aggaagaacc ttgctggggt ttcttctgca acgtctcagt cccgtgtggg ttatctggag 1140
cagacagaga tttaccaaaag cacgcctcct ggacttcagg ctgcgctgg caggctcgtg 1200
gctgcaaaat caactttggc agcaagagt gatgctacta gaggggatcc gttagggata 1260
agtggaaaag ctttcaggga ggagatccgt aagaagattg agaaatggca agaacctcct 1320
cctgcaagac agcctaagcc acttctgtt cctgattctg aaccgaagaa agaaggggt 1380
ggtcgccgtc taagaaaaat gaaagaaagg tagccttttt cactctactt tgtgtcetta 1440
attactgtag attgagttct attcacctgt atttattttg ttgcattctt acgtttctct 1500
ttaaatcagg tatcaagtaa cagatatgag gaagctggcc aacagaatgg cgtttggtac 1560
acctgaagag agctccctcg gtaatatatc ttgtagttac acttgtaaat ggccacttat 1620

```

```

aaggcactta gtctaatac tactcttcat gatgataggat gatggactag gagaaggtta 1680
tggaatgctt ggccaggcag gaagcaacag gctgcgagta tccagtgttc cgagcaagct 1740
taagattaat gctaaggctc ccaaaaagta agtgttcctc tatttctcct gtgttttttc 1800
ggatttatca tgtaaatatt tttactctta caaattatcc tgcctgttc ttcttccatc 1860
atatctcatt tgcgtcttta tatcaattac tttttcagge ttaaagaaag gcagtatgcy 1920
ggtggtgcyg ctacctctgg tttgacatcg agcctggctt tcactcctgt gcaggtacaa 1980
acatttcatt cgattcttga caaaagttg atcctgtgtt ccatttgcac cactgtctga 2040
ctccaattgg ttatctattt gacaggggat agagttgtgc aatcctcagc aggctttagg 2100
attaggaagt gggactcaa gcacttactt ctcagagtca ggaaccttct cgaagctgaa 2160
gaagatctaa
2170

```

```

<210> 15
<211> 20
<212> DNA
<213> 人工序列

```

```

<220>
<223> 人工序列的描述:
寡核苷酸

```

```

<400> 15
accttaggcg acttttgaac
20

```

```

<210> 16
<211> 24
<212> DNA
<213> 人工序列

```

```

<220>
<223> 人工序列的描述:
寡核苷酸

```

```

<400> 16
aaacgcttac catatctctt tcta
24

```

```

<210> 17
<211> 113
<212> DNA
<213> 鼠耳芥

```

```

<400> 17
aaacactagt cgctcgctgc tcttcaattt tcttctcgaa tctaactgat tgatttctcc 60
ttcgattctt caggagaatc actgaagctt ttgcctccca agtagaaaga gat 113

```

```

<210> 18
<211> 218
<212> DNA
<213> 鼠耳芥

```

```

<400> 18
aatatggaag acagagatnc aagtcttgaa aagccgagca ctaaaagtgt aaaaatgaac 60

```

caaaggtgga aagaaactgc tttctctatc tcatgtctgt ttttaaggttt cttcgggtcac 120
 ttaagagaca aaaggcattg ttttgatcac tctttggaaa cgttttataa attttatttt 180
 tgtattagag ccaaaaaaaaa aaaaaaaaaa aaaaaaaaa 218

<210> 19

<211> 4140

<212> DNA

<213> 鼠耳芥

<400> 19

cagtacactt agetacactg gatccaagtc tagtgctaaa ctcaaacctc gtggtttttag 60
 accaaaaatct cttctctctc gtttctctct tctctcatcat atctttcatc ttctccacca 120
 gaatttgttt taggctctcc ttctctctgtt tctttttctc ccaaagaaac aattagatat 180
 ggagaacctt accctagttt cttgctcagc ttctctctcca aagctgtaa ttggatgcaa 240
 tttcacttcc tcgctgaaaa accctactgg gttttctcgt cggactccta atattgtcct 300
 ccggtgttcc aaaatatctg cctctgctca atctcaatct ccctcttcgc gtccggagaa 360
 cactggagaa atcgggttagt ttgcaaattc cactcgacac tctattatag caaatgcca 420
 aattttccgg aaaaatttcc agtttattac ttttatctat cttattgaaa ctcaaattgc 480
 gaacctttt cgactggttt aatatgagct tatgaattgc tatactctct aaaaaatcc 540
 acactttgtg aatttgcaat ttgaattctt gtagaaacca ttcattgtta gaattgtta 600
 ctttaagttt atgttcgatt tgcagtggtt gtgaaacaga gaagcaaagc ttttgcaagt 660
 atattttctt cgagtcgtga tcaacagaca acttctgttg cttcccctag tgtgcctgtg 720
 ccaccaccat cttcatcaac catgtaattt tcttggtttt ggacaatgtg cttagtttgt 780
 atgtcgtttg attcttggtt attaaattgt gttttttctt ttttcttgta gaggatcacc 840
 acttttctgg attgggtgtg gtgttggtct atcagctttg ttctcatatg tgagtatcaa 900
 gattccttcc taatttttt ttctctata aatattctt cttgcttcaa tattgattaa 960
 taagtgtttg acctttttt ttttctgatg gcattgcagg taacttcaa tttaaaggta 1020
 cagatacttg gccctctggt tttacgggac ttttgttctc tagtctgttg cagaaccacg 1080
 attttatgct tcatgtcaac tctagtgtat tgtgctcatg tatctgagat agttttattc 1140
 actaaactgg ttatcttaac aagggtgaact gtttgctcac acttgttgaa ccgtttatat 1200
 aagcatcgaa cttttgcctc tctttttttg ggtagtcact tgattcgtag atggtaacct 1260
 acataccatt atggttttag tgatgcaact caggtattca gacttatagt cattttcgca 1320
 actccagtat ttgattgaaa tatattatac aagttgtcat tgctttctct cattattctc 1380
 taaccggctg ttactctctt tggattttt tttttgctt ggttttagaaa tatgcaatgc 1440
 aaacagctat gaagacgatg atgaaccaa tgaatacgca aaatagccag ttaataaatt 1500
 ctggattccc atcaggatca ccttttccgt ttccatttcc tctcaaaca agtctgtct 1560
 cctcgccatt ccaatctcaa tcccagtctt caggtgctac cgttgatgtg acagcgacaa 1620
 aagtagagac acctccttca actaaaccga aacctacacc tgcaaaggat atagaggtgg 1680
 ataagccaag tgttgtctta gaggcaagca aagagaagaa agaagaaaag aactatggta 1740
 gattctttt ctgtttcaga aatcaacgtc ttttcatttg tattctcaat tttgactttc 1800
 ttctttctc attttccaag cttctaactt ggaagctgat ttacttttgg atgcagcctt 1860
 tgaagacatt tcacccgagg aaaccacaaa agaaagccca tttagcaact atgcagaagt 1920
 ctctgaaact aattccccca aagaaactcg cttgtttgag gatgtaagt tcgttttctt 1980
 ttgtatttcc acagcacacc aagtggatgat ttaaaaactg gacatagttt tgctaacctt 2040
 ctatgctctc ttattgatct ctgggtgaag gtcttgcaaa atggagctgg tccggcaaat 2100
 ggtgccactg cttcagaggt ttttcaatct ttgggtgagt tattgaattt cagttttcat 2160
 cactatcagc gcactgtgca tgattcatga ttaaggctac ggatttcaat tttattttat 2220
 agcatatgcc aacaattata acaaaggaa gatatgaaat tggtgataaa gaggaatgag 2280
 ttggcttcaa aaggatctac tccgttactt ttgtcctct gctagtcgtt gatctgtatt 2340
 ggtataacca tataagactt gcaggatatt accttggcaa tctgtttcat atctcatgtg 2400
 ttatgattct tttttcttat atgctcacgt tattgtctct cttttcttca ttctaaattt 2460
 aaaactgaat cctgagctg tctattgttt acacaggtgg tgggaaagga gggccgggtt 2520
 tatctgtaga agcttttagag aaaatgatgg aagatccaac agtccagaag atggtttacc 2580
 cgttaactcat cttcccctag acattgtctt taaatgcatc cattaagttt atctttaaaa 2640

```

ctggttgctt agtggacatt tggtaacatt gcatgtataa atgcagatac ttgcctgagg 2700
agatgaggaa cccagaaact ttcaaatgta agtcttttaa tatttaatcc tgctatcatt 2760
cttttattag tcttcatttt tacatatttc taaagactaa aggttacatg actagctttt 2820
gaatgatgta attcgtttat aggttgatcc aatggttatc taaatttaa atacagtttg 2880
gtacttattg tctccgcttg gaattttgta gggatgctta aaaatcctca gtaccgtcaa 2940
caactacagg acatggtgta agagctccat ttacgaaca atttagttgt ttccattgct 3000
tttaagaatg tctaaactat gtaattaaga aatactcttg tttgtttctt ttcattgaatt 3060
taggaataat atgagtggga gtggtgaatg ggacaagcga atgacagata cattgaagaa 3120
ttttgacctg aatagtcctg aagtgaagca acaattcagt aagacaaatc tcagtttgta 3180
ccaagttaat agtacgttaa ataggtctga tactcaatga ttgaatctgt atttgtcaga 3240
tcaaatagga ctaactccag aagaagtc atctaagatc atggagaacc ctgatgttgc 3300
catggcattc cagaatccta gagtccaagc agcgtaatg gaagtacgtt ttcttttaac 3360
ctgaataaga gaattgctta attttaccctt acttctttct tcatacaaaa cagaaaccaa 3420
ttacattctt gttggtggtg cagtgtcag agaaccctat gaacatcatg aagtaccaa 3480
acgacaaaga ggtaataata ctgccacttc tccattgccc aaaaaggcga ttactttttt 3540
aagaaatttg aggttattat acattgattg caggtaatgg atgtgttcaa caagatatcg 3600
cagctcttcc caggaatgac gggttgaaaa agctcacgctc tttggttcta tcaaaaatgt 3660
cacattgtct ttagcttttt gttagggagaa aaaaatgttt ttttttttgc aaagagtctt 3720
cagttttggt cagatcagag aattgtgtac catgtaatc ttaaaccgcg tcgggaattg 3780
gagtcgtgtg aaaacgcccg tgcgtgtggt tggatgaat attatacaat agaatttgtt 3840
gtcttaccaa aaaaagtcta tgaagacact gaagagcaaa ttattatttt taagggaaaa 3900
tttccaaaat aaacttcatg tattcaaat ttgcttgaaa aaacctcaat ttttttgtt 3960
tgagattgtg tgaataaatc tgccaatatt ttgttttagc aatttaaaaa attgaagttt 4020
ttttctcgca aattttaaat agttgtgatt tattttggaa ttttacctta ttttaatat 4080
ccaaaaggag aagtgacgtg gcgatatcga agcggtttaa tgaagtgat gccccatctt 4140

```

<210> 20
<211> 77
<212> DNA
<213> 鼠耳芥

```

<400> 20
ccacggtcc gctccaccag aatttgttt aggtctctct tcttctggtt cttttctcc 60
caaagaaca attagat 77

```

<210> 21
<211> 354
<212> DNA
<213> 鼠耳芥

```

<400> 21
aaaagctcac gtctttggtt ctatcaaaa tgtcacattg tctttagctt tttgtaggga 60
gaaaaaatg ttttttttt tgcaaagagt cttcagtttt ggtcagatca gagaattgtg 120
taccatgta atcttaaacy cggtcgggaa ttggagtcgt gtgaaaacgc cgctgctggt 180
gtttggtatg aatattatac aatagaattt gttgtcttac caaaaaaagt ctatgaagac 240
actgaagagc aaattattat ttttaagggg aaatttccaa aataaacttc atgtattcaa 300
aatttgcttg aaaaaacctc aatttttttt gttgaaaaaa aaaaaaaaaa aaaa 354

```

<210> 22
<211> 24
<212> DNA
<213> 人工序列

<220>

<223> 人工序列的描述:

寡核苷酸

<400> 22

cagaccacaa taccttcaaa aata

24

<210> 23

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:

寡核苷酸

<400> 23

ccattgtgtc tccctcccgc tgtt

24

<210> 24

<211> 5077

<212> DNA

<213> 鼠耳芥

<400> 24

```

atgattgtaa aacttgacag ggaggattht agtgatacac ttcgagtact tgttgcaact 60
gattgccact tgggctacat ggagaaggat gaaattagge ggcattgattc atttaaggct 120
ttcgaagaga tatgttctat agctgaggag aaacaggctt ggtattcagt atctatccct 180
tgccagtatt atcttgcggt tgaatcatct aacatattat cttaaataaa aatcttctcc 240
caatattatg agtagtaaac agtgttctac ctaattttta caaaaattca accaattgcy 300
aggaagaatt ctcaaaaagt ttcatatctt cttttttcac tcttttgaaa caggtggact 360
tcttactcct cggagggtgat ctttttcatg agaataaacc ctctagaact acgtagttag 420
aagccattga aattcttctg cggcactgtc tgaatgataa accagtgagc tttcaagtag 480
tcagcgacca gacagtaaat tttcagaatg cgtgagactc tatectttct gctattaatc 540
taatcataac aggaataaat ttcaactgaa ctaattaatt ggcaaattgg ctcaaattcg 600
tgtatagatc tacgtattct tattaatecc ttgacattat tttctggcta caggtttggt 660
caagtcaatt acgaggatcc aacttcaat gtaggcttgc ccgtgttcag tattcatgga 720
aaccatgatg atccagccgg agtggtagat cacttacatc tgcattgctc tgttatgcaa 780
actcatttga ataggtatat agaactggat tagttagtga ataggtattt tattgtggtt 840
ttgttctatg tctcttatgg ctacaggaca atctttctgc aattgatatt ctttccgcat 900
gcaaccttgt gaactattht ggaaagatgg ttcttgggtg ttctgggtgt ggccagatta 960
ctctctacc tatacttatg aagaaggttg gtgtaaagaa tttctaacct agacacctgg 1020
ctccccctga cttcttggac tatcatttaa tcaaattaat gtttagggct caacaaccgt 1080
ggctctctat ggtttaggaa acatcagggg tgaacgtctc aatagaatgt ttcaggtaat 1140
ccagaggacc ctacacttht gctatacaat tghtaattgt gttaatattt attggtttca 1200
cagaccccac atgctgtcca atggatgagg cctgaagttc aagaaggatg tgatgtttct 1260
gactggttca acattctggt gcttcatcaa aataggttga ttccattgct ataacatctt 1320
ttagatcggt ttcttactca ttctgtatca gaaaatttga tactgtatc atatgacttg 1380
caggttgaaa tcaaaccaca aaaatgcaat aagtgagcac tttcttccac gtttccctga 1440
cttcattgtg tggggccatg agcatgaatg cctaactgac ccccaggctc atgaaaaatt 1500
tgatttttgg agttattgca tttaaataag agtgagccac aatgttactt gcctctttga 1560
gctaaaagct attaaactth tgaaggagggt atctggaatg ggcttccaca tcacacaacc 1620

```



```

aggatcttct gtggcaacat cacttattga tggggaatcg aagccaaaac atgttcttct 1680
cttagaaatc aaggttcttc agcaacaat ctgaaatttc atcttcactt tattcgtact 1740
tcattttctg gtcttttttc ctccctttca atcaagcatg taagcttgag tgacttaaaa 1800
tatatgactt acagggaaat caataticgc ctacgaagat acctttgaca tctgtgagge 1860
cttttgagta tacagaggta aagtttactt ttccttaata tgttatggtg gtggcagact 1920
tctttgctta catattttca aagtgcagat tgttttaaag gatgaaagtg atattgatcc 1980
caatgatcaa aactcaattc tggaaactt ggataaagtg gtacctattc cctcttctca 2040
tagttcatgt ggatatcttt tctcctgccc tttttgaata accagtcact gaatgtctct 2100
actaatatct acaaaattgt taggtcagaa atctaataga gaaagctagc aaaaaagctg 2160
ttaacagatc agagatcaaa ctcccattgg ttcgaatcaa ggtaacttgt ttccaagttt 2220
tcttcaaact gctgcaaatt cttagcaaac tcatataatt aaacctttat tttctaacc 2280
aactctagag gctaggcttt gccagtttga tgcattgcaca cccatagcca caaacagata 2340
attgttatta agaataatga atgactgaca aaagactaag atctgcttca tctttcaggt 2400
agattattct ggatttatga cgataaatcc tcaaagattt ggacagaaat atgtgggaaa 2460
ggtacctaga aattagttac tgaacatga tggtcaccat acttctttga atgttggtta 2520
actaatgaca aagtcccaaa cacttacagg ttgcaaatcc ccaggacatt ttgatatttt 2580
ccaaggcttc taagaagggt cggagcgaag gtaagggcat tgggtgacta gtaatttata 2640
caattttggt tggattagat tgatgcacgt gcttttactc taacttgtaa tagcttatct 2700
ggcaaaaatt acggttaagt agtgtatctg agatatagta atgtagaaca atatgggctt 2760
atgataacct cctttgttgt tttattgtcg gtattataat tctcgtcata tatatcatga 2820
ctactaactt tctgttgtgt ggagcttgat attgatgat tgagtgttaa ttttctttct 2880
gttccacttt tcttgttata gttcatgttt ctctgtgtgt aacctatagc atcaaaattt 2940
tgccaatctt atggattatc tctagttagt atatatggga aatttgccat tttgataatt 3000
ttttgtctta gtgaattgaa tggcaatgat gcatgtcctg atggttgtcc agtgatccag 3060
ttatgatata tttcaatctt ccatttcaca gccaacatcg atgattctga gcggcttcgt 3120
ccagaagaac tgaaccagca gaatatagaa gctttagtag ctgaaagcaa cctggtacat 3180
cctgcaacct tctttcctta tgattgtgtt attatcgtca acccctgtag aactttgcca 3240
cagaatgata tagacttggg tagttaccaa atgggcatga gtacactatg ggatgatcat 3300
tctattttct tccgcagaaa atggagatcc ttccagtaa cgatctggat gttgctcttc 3360
acaattttgt gaacaaggat gataaactag ccttctactc atgctgtcag tacaatcttc 3420
aagagactcg tgtatgtact attttttact tcaccattca atacaaagt ctgcatagga 3480
tattattttt atttcgtagc acgtccttgt tattgctttt atgatttatc tcttccctct 3540
ttttgtacag ggtaaacttg caaaggattc agatgccaaag aaatttgagg aagatgactt 3600
gattcttaaa gtgggagagt gcttagaggc aagaagatat agattcagtt agttctgccg 3660
cagattatga gaaccagcag aatattgatc tcaactgcat tattgttctg gcaggaacgc 3720
ttgaaagata ggtccactcg acccactggt tctcaccagt tttatccac tggattgact 3780
tcagagggtt aaattctctt ttttagattt tcttgctc tgctctccg ttggtttctc 3840
acagtgctat tttctacctg agattggtac agaatttgac aaaaggaagc agtggcatcg 3900
cgaatgcttc gttcagtgat gatgaagaca caactcagat gtctggttta gctcctcca 3960
ctagaggacg aagaggttca tccactgcta atacaactcg tggtagagct aaagcccaa 4020
ccagaggacg aggcctggtt aaggcctcaa gtgcgatgaa gcaaaccact cttgatagtt 4080
ctcttggttt ccgccagtct caaaggtaac tttttgacag cacatttaac cagttaggg 4140
taggattcac ggacgtgcaa ggaaatgatt ggcattcacta gctagctaat gttatgtccc 4200
taatttgctt ttcatagatc tgcttcggct gctgcttcag ctgccttcaa aagtgtctcc 4260
accattggag aagatgatgt agattctcct tcaagcgaag aagtcgagcc tgaagatttt 4320
aacaacctg acagcagttc ggtatggact attccttaca ctgttattca tttgttact 4380
accataagaa agcccatgta aaaacttgac aacatataac ttttggcatt cttatttctc 4440
tatttgaagt aaattttgcg tttttacttt tcttgattct tgtttgatc cactaaagg 4500
aggacgatga gagcactaaa ggcaaaggac gtaaaagacc agctactact aagagaggca 4560
gaggtagagg ttctgggact tcaaaacgtg gttagaaaaa cgaaagctct tcttactta 4620
ataggctact cagtacaaa gacgatgacg aggacgaaga tgatgaagac agagaaaaga 4680
agcttaacaa atctcagcct cgggtttggt aatcacatct attttccctt cttctgctgc 4740
ttattagcag gttttagtaa gttgttgta accatttgag atcaaagctc acttaatagt 4800
acaattttaa tatgcagggt acaaggaact atggagctct aagaagataa atacatatca 4860
aaccccaatc tctgacatca caacgaagct tcatttttct gttattttct agcgacctct 4920

```

caagcggaaac aacttctgaa gaagagaaat tagtactaac aagagttctg tgagatgatg 4980
 tacagagaat tttgtagtgt ttttttttct tgctcttttt aaggttacgt tgttgatgaa 5040
 tgaggcaata tgattaacgt cagtaagaag tctaaaaa 5077

<210> 25
 <211> 18
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列的描述:
 寡核苷酸

<400> 25
 tgtaaacga cggccagt 18

<210> 26
 <211> 255
 <212> DNA
 <213> 鼠耳芥

<400> 26
 atacatatca aacccaatc tctgacatca caacgaagct tcatttttct gttattttct 60
 agcgacctct caagcggaaac aacttctgaa gaagagaaat tagtactaac aagagttctg 120
 tgagatgatg tacagagaat tttgtagtgt ttttttttct tgctcttttt aaggttacgt 180
 tgttgatgaa tgaggcaata tgattaacgt cagtaagaag tctaaaaaaaa aaaaaaaaaa 240
 aaaaaaaaaa aaaaaa 255

<210> 27
 <211> 2935
 <212> DNA
 <213> 鼠耳芥

<400> 27
 tcatgcatga actggcctag caccagaaga aagctgcaca ttcgcggcat attcacgtgc 60
 ccacaagtca tagtgaacaa tctcttcaag agacgggtgat gttaccaact cgtttcgatg 120
 tttatcgcac gtttaattcca caaccttgaa gatatccaaa tagcttatcc tgtaaacaaa 180
 agtgagaata taaacaattg tgattcgtat caagaacttc attgagatgc tcaaaaactga 240
 aaaataattc ttacttttca tcaatgaaca tttcaacagc tttctcattg gcggcgctga 300
 gaactccagt cattgtgcct ccagctcgtc cagcagcata agcaagatcc atggatgggt 360
 atttcacatt gtctggtttc ttgaaagtca atgaaccgag tctgccaaaa tccacaattg 420
 taaacaactt ttggtttttag gtgctgaatg ctgatagata aggcagtggg cctaaccag 480
 ttttaactgat ccacacaaa acagtagcaa aataaccaat tgcaaaaacca aaccgaagac 540
 cgattcgggt tcatttttta tcttatctaa acaacctaaa accaaaactga aaacaagatt 600
 ggggaacttt tcttgggtgat aattaaatt ttcaactaag cttagcttca cacttgataa 660
 acagagagta tataaatgtg gttagcttac ttgcaaaggc caagtcttgg ccaagttact 720
 tcagaacaag gaactctatc gggccatgac atggtgtaga gaatcggtaa acgcatatca 780
 ggccaacca attgagcaag cacagatgaa tcctgtggaa caaaacaaat acatgttata 840
 cagttatfff tttaaaaccg gaaaaataat aathtagtta gtaatgtttc agcaagacct 900
 gtgtttcaat catggaatgt atgatacttt gcggatgaat gacaatctct atatcgatca 960
 actcagctcc aaacaataa tgcgcttcaa tgacctcaag accctgtttc aaaaaatcaa 1020
 gaactcatct acctgatca aaggtatfff caaaatcaga gtttaacctt aggagaaaat 1080

```

aatcttaacc ttgttgaaaa gcgtagcaga gtccacagtg attttcttc ccatgttcca 1140
gtttggatgc ttcaacgcat ccgctacttt aacttccttt agcttttcga caggccaatc 1200
cctttttcaa aatccagtga aaagtttcca ttaaccaaac gagaattgag aagaaaaaaa 1260
gtctatgcag agagagaaga atatcgaaac aaacctaaaa gctccaccag atgcagtcaa 1320
gattatcttg cgcagagcgc cttcaggcaa accttgaata cactagagaa cataaaagaa 1380
gatttttcac tcaaattgcc agaggttgaa cttgcattaa gaccaacgct gaactcaata 1440
tgaaagttga ggtacttaat tctatgtgat ttgtgatacc tgaaatatgg cagaatgttc 1500
tgaatctgcc ggaagaatct ttacattatg tttgttggca agcgggaagca cgaaaggacc 1560
acctgcgatt aatgtctctt tgtttgcaag agcaatgtcc tttcctgctt caattgcagc 1620
aacgtaggc tgcagtaaaa ataagcaaca agctttatca tctgcaactt tcttttttca 1680
tatcctctta ataaggttta ataacaaaaa attagagtat atacctttag tccccacaaa 1740
cctactattc cggtacaacac ggttacagct tcaggatgtc gggcaacctg ttgatgaaca 1800
taataagtaa aaacctatct aactacaat caaaactaac aatgaacta acctcaatca 1860
ctccttgctc tcctggaata atctcgagtt tatagtccaa atcagctaaa gcctctttaa 1920
gctcattaat cagtgactcg tttctaacag caaccaatgc aggttaaat ctccttacct 1980
gccaccattc aaaatagaat cacagaacca tactatagag atttcttgag attgcagaag 2040
caaaagccta aaccagaacc tgatttctct ggtttgatct gatacataac gagttaatac 2100
tatcttgctt atgatactac cactgaactg agaattaaac tgaattccaa gtggctctgaa 2160
tgacaaattg gagagactca atactaattt ttttacaaat gaagccaact tacctgatca 2220
gcaagtagag taacattcga accagcagct agagccacaa ctctgaattt gtcaggattc 2280
tcagccacaa tatccaatgt ctgcaaaatg gaagttcttg tcgataaaaa tgatgcaaca 2340
ataactcagt aagaaaaaaaa tatcattctt ctatgagtct agtcattcat aagacaaact 2400
taaagtctgg tcatactcaa gaactgcaca ataatgcctt aatcgaaata aaacctgagt 2460
gccaatagaa ccagtagatc caacgataga gatgggtttt ggtccatccc aagattgacg 2520
aggcgccctc gggacagctc tcccaggcca tgctggagga ggttggtgtt gctgctgcac 2580
tttactgaa cacttaacac cttttccaaa acctctccct tgattcctcc tcctcaaaact 2640
aaaccacct gtgaaacact ccaaagatgt aaaatttaaa actctacgac ctaaagcaaa 2700
caaaaaaaaa tcgaattgaa gaaataacag attacctaga tagagaaatt cacaagagcc 2760
taagacaact aatgaaagtt tgcaacttta atcgaaaaga gagttgacca aggaggagga 2820
aagaagagag gaagaagaag aaacctgaga gtttagggat tggattgaac ctggagggtat 2880
ccaagaaaga aatagctttg gattcagctg gagatagtga gtttaatgtc atcat 2935

```

<210> 28
 <211> 1434
 <212> DNA
 <213> 鼠耳芥

<220>
 <221> CDS
 <222> (1)..(1434)

```

<400> 28
atg atg aca tta aac tca cta tct cca gct gaa tcc aaa gct att tct 48
Met Met Thr Leu Asn Ser Leu Ser Pro Ala Glu Ser Lys Ala Ile Ser
  1 5 10 15
ttc ttg gat acc tcc agg ttc aat cca atc cct aaa ctg tca ggt ggg 96
Phe Leu Asp Thr Ser Arg Phe Asn Pro Ile Pro Lys Leu Ser Gly Gly
  20 25 30
ttt agt ttg agg agg agg rat caa ggg aga ggt ttt gga aaa ggt gtt 144
Phe Ser Leu Arg Arg Arg Xaa Gln Gly Arg Gly Phe Gly Lys Gly Val
  35 40 45

```

aag tgt tca gtg aaa gtg cag cag caa caa caa cct cct cca gca tgg 192
Lys Cys Ser Val Lys Val Gln Gln Gln Gln Gln Pro Pro Pro Ala Trp
50 55 60

cct ggg aga gct gty cct gag gcg cct cgt caa tct tgg gat gga cca 240
Pro Gly Arg Ala Xaa Pro Glu Ala Pro Arg Gln Ser Trp Asp Gly Pro
65 70 75 80

aaa ccc atc tct atc gtt gga tct act ggt tcy aty ggc act cag aca 288
Lys Pro Ile Ser Ile Val Gly Ser Thr Gly Xaa Xaa Gly Thr Gln Thr
85 90 95

ttg gat att gtg gct gag aat cct gac aaa tty aga gtt gtg gct cta 336
Leu Asp Ile Val Ala Glu Asn Pro Asp Lys Phe Arg Val Val Ala Leu
100 105 110

gct gct ggt tcg aat gtt act cta ctt gct gat cag gta agg aga ttt 384
Ala Ala Gly Ser Asn Val Thr Leu Leu Ala Asp Gln Val Arg Arg Phe
115 120 125

aag cct gcr ttg gtt gct gtt aga aac gag tca ctg att aat gag ctt 432
Lys Pro Xaa Leu Val Ala Val Arg Asn Glu Ser Leu Ile Asn Glu Leu
130 135 140

aaa gag gct tta gct gat ttg gac tat aaa cyc gag att att cca gga 480
Lys Glu Ala Leu Ala Asp Leu Asp Tyr Lys Xaa Glu Ile Ile Pro Gly
145 150 155 160

gag cwa gga gtg att gag gtt gcc cga cat cct gaa gct gta acc gtt 528
Glu Xaa Gly Val Ile Glu Val Ala Arg His Pro Glu Ala Val Thr Val
165 170 175

gtt acc gga ata gta ggt tgt gcg gga ctg mag cct acg gtt gct gca 576
Val Thr Gly Ile Val Gly Cys Ala Gly Leu Xaa Pro Thr Val Ala Ala
180 185 190

att gaa gca gga aag gac att gct ctt gca aac aaa gag aca tta atc 624
Ile Glu Ala Gly Lys Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile
195 200 205

gca ggt ggt cct ttc gtg ctt ccg ctt gcc aac aaa cat aat gta aag 672
Ala Gly Gly Pro Phe Val Leu Pro Leu Ala Asn Lys His Asn Val Lys
210 215 220

att ctt ccg gca gat tca gaa cat tct gcc ata ttt cag tgt att caa 720
Ile Leu Pro Ala Asp Ser Glu His Ser Ala Ile Phe Gln Cys Ile Gln
225 230 235 240

ggt ttg cct gaa ggc gct ctg gcg aag ata atc ttg act gca tct ggt 768
Gly Leu Pro Glu Gly Ala Leu Arg Lys Ile Ile Leu Thr Ala Ser Gly
245 250 255

gga gct ttt agg gat tgg cct gtc gaa aag cta aag gaa gtt aaa gta 816
Gly Ala Phe Arg Asp Trp Pro Val Glu Lys Leu Lys Glu Val Lys Val
260 265 270

gcg gat gcg ttg aag cat cca aac tgg aac atg gga aag aaa atc act 864
 Ala Asp Ala Leu Lys His Pro Asn Trp Asn Met Gly Lys Lys Ile Thr
 275 280 285

gtg gac tct gct acg ctt ttc aac aag ggt ctt gag gtc att gaa gcg 912
 Val Asp Ser Ala Thr Leu Phe Asn Lys Gly Leu Glu Val Ile Glu Ala
 290 295 300

cat tat ttg ttt gga gct gag tat gac gat ata gag att gtc att cat 960
 His Tyr Leu Phe Gly Ala Glu Tyr Asp Asp Ile Glu Ile Val Ile His
 305 310 315 320

cck caa agt atc ata cat tcc atg att gaa aca cag gat tca tct gtg 1008
 Xaa Gln Ser Ile Ile His Ser Met Ile Glu Thr Gln Asp Ser Ser Val
 325 330 335

ctt gct caa ttg ggt tgg cct gat atg cgt tta ccg att ctc tac acc 1056
 Leu Ala Gln Leu Gly Trp Pro Asp Met Arg Leu Pro Ile Leu Tyr Thr
 340 345 350

atg tca tgg ccc gat aga gtt cct tgt tct gaa gta act tgg ccw aga 1104
 Met Ser Trp Pro Asp Arg Val Pro Cys Ser Glu Val Thr Trp Xaa Arg
 355 360 365

ctt gac ctt tgc aaa ctc ggt tca ttg act ttc aag aaa cca gac aat 1152
 Leu Asp Leu Cys Lys Leu Gly Ser Leu Thr Phe Lys Lys Pro Asp Asn
 370 375 380

gtg aaa tac cca tcc atg gat ctt gct tat gct gct gga cga gct gga 1200
 Val Lys Tyr Pro Ser Met Asp Leu Ala Tyr Ala Ala Gly Arg Ala Gly
 385 390 395 400

ggc aca atg act gga gtt ctc agc gcc gcc aat gag aaa gct gtt gaa 1248
 Gly Thr Met Thr Gly Val Leu Ser Ala Ala Asn Glu Lys Ala Val Glu
 405 410 415

atg tty att gat gaa aag ata agc tat ttg gat atc ttc aag gtt gtg 1296
 Met Phe Ile Asp Glu Lys Ile Ser Tyr Leu Asp Ile Phe Lys Val Val
 420 425 430

gaa tta aca tgc gat aaa cat cga aac gag ttg gta aca tca ccg tct 1344
 Glu Leu Thr Cys Asp Lys His Arg Asn Glu Leu Val Thr Ser Pro Ser
 435 440 445

ctt gaa gag att gtt cac tat gac ttg tgg gca cgt gaa tat gcc gcg 1392
 Leu Glu Glu Ile Val His Tyr Asp Leu Trp Ala Arg Glu Tyr Ala Ala
 450 455 460

rat gtg cag ctt tct tct ggt gct agg cca gtt cat gca tga 1434
 Xaa Val Gln Leu Ser Ser Gly Ala Arg Pro Val His Ala
 465 470 475

<210> 29

<211> 477
 <212> PRT
 <213> 鼠耳芥

<400> 29

Met Met Thr Leu Asn Ser Leu Ser Pro Ala Glu Ser Lys Ala Ile Ser
 1 5 10 15
 Phe Leu Asp Thr Ser Arg Phe Asn Pro Ile Pro Lys Leu Ser Gly Gly
 20 25 30
 Phe Ser Leu Arg Arg Arg Xaa Gln Gly Arg Gly Phe Gly Lys Gly Val
 35 40 45
 Lys Cys Ser Val Lys Val Gln Gln Gln Gln Gln Pro Pro Pro Ala Trp
 50 55 60
 Pro Gly Arg Ala Xaa Pro Glu Ala Pro Arg Gln Ser Trp Asp Gly Pro
 65 70 75 80
 Lys Pro Ile Ser Ile Val Gly Ser Thr Gly Xaa Xaa Gly Thr Gln Thr
 85 90 95
 Leu Asp Ile Val Ala Glu Asn Pro Asp Lys Phe Arg Val Val Ala Leu
 100 105 110
 Ala Ala Gly Ser Asn Val Thr Leu Leu Ala Asp Gln Val Arg Arg Phe
 115 120 125
 Lys Pro Xaa Leu Val Ala Val Arg Asn Glu Ser Leu Ile Asn Glu Leu
 130 135 140
 Lys Glu Ala Leu Ala Asp Leu Asp Tyr Lys Xaa Glu Ile Ile Pro Gly
 145 150 155 160
 Glu Xaa Gly Val Ile Glu Val Ala Arg His Pro Glu Ala Val Thr Val
 165 170 175
 Val Thr Gly Ile Val Gly Cys Ala Gly Leu Xaa Pro Thr Val Ala Ala
 180 185 190
 Ile Glu Ala Gly Lys Asp Ile Ala Leu Ala Asn Lys Glu Thr Leu Ile
 195 200 205
 Ala Gly Gly Pro Phe Val Leu Pro Leu Ala Asn Lys His Asn Val Lys
 210 215 220
 Ile Leu Pro Ala Asp Ser Glu His Ser Ala Ile Phe Gln Cys Ile Gln
 225 230 235 240
 Gly Leu Pro Glu Gly Ala Leu Arg Lys Ile Ile Leu Thr Ala Ser Gly
 245 250 255
 Gly Ala Phe Arg Asp Trp Pro Val Glu Lys Leu Lys Glu Val Lys Val
 260 265 270
 Ala Asp Ala Leu Lys His Pro Asn Trp Asn Met Gly Lys Lys Ile Thr
 275 280 285
 Val Asp Ser Ala Thr Leu Phe Asn Lys Gly Leu Glu Val Ile Glu Ala
 290 295 300
 His Tyr Leu Phe Gly Ala Glu Tyr Asp Asp Ile Glu Ile Val Ile His
 305 310 315 320
 Xaa Gln Ser Ile Ile His Ser Met Ile Glu Thr Gln Asp Ser Ser Val
 325 330 335
 Leu Ala Gln Leu Gly Trp Pro Asp Met Arg Leu Pro Ile Leu Tyr Thr
 340 345 350
 Met Ser Trp Pro Asp Arg Val Pro Cys Ser Glu Val Thr Trp Xaa Arg
 355 360 365
 Leu Asp Leu Cys Lys Leu Gly Ser Leu Thr Phe Lys Lys Pro Asp Asn
 370 375 380
 Val Lys Tyr Pro Ser Met Asp Leu Ala Tyr Ala Ala Gly Arg Ala Gly
 385 390 395 400

Gly Thr Met Thr Gly Val Leu Ser Ala Ala Asn Glu Lys Ala Val Glu
 405 410 415
 Met Phe Ile Asp Glu Lys Ile Ser Tyr Leu Asp Ile Phe Lys Val Val
 420 425 430
 Glu Leu Thr Cys Asp Lys His Arg Asn Glu Leu Val Thr Ser Pro Ser
 435 440 445
 Leu Glu Glu Ile Val His Tyr Asp Leu Trp Ala Arg Glu Tyr Ala Ala
 450 455 460
 Xaa Val Gln Leu Ser Ser Gly Ala Arg Pro Val His Ala
 465 470 475