

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 00808457.2

[51] Int. Cl.

C12N 15/53 (2006.01)

C12N 15/82 (2006.01)

A01H 5/00 (2006.01)

A01N 65/00 (2006.01)

[45] 授权公告日 2008 年 5 月 14 日

[11] 授权公告号 CN 100387721C

[22] 申请日 2000.4.14 [21] 申请号 00808457.2

[30] 优先权

[32] 1999.4.15 [33] US [31] 60/129,899

[32] 1999.7.30 [33] US [31] 60/146,461

[86] 国际申请 PCT/US2000/010367 2000.4.14

[87] 国际公布 WO2000/063389 英 2000.10.26

[85] 进入国家阶段日期 2001.12.4

[73] 专利权人 卡尔根尼有限公司

地址 美国加利福尼亚州

[72] 发明人 G·M·基肖尔 A·博罗纳

N·坎波斯

[56] 参考文献

EP0723017A 1996.7.24

WO9911757A 1999.3.11

审查员 曾繁辉

[74] 专利代理机构 北京纪凯知识产权代理有限公司

代理人 程伟

权利要求书 3 页 说明书 32 页 附图 2 页

[54] 发明名称

编码参与类异戊二烯合成的蛋白质的核酸序列

[57] 摘要

本发明提供了产生类异戊二烯化合物含量及组成改变了的植物和种子的方法。这些方法尤其用于提高植物中的类异戊二烯水平，以及在宿主植物细胞中提供所希望的类异戊二烯组成。

1.一种编码源自真核细胞的具有序列 SEQ ID NO: 2 的 1-脱氧-D-木酮糖 5-磷酸还原异构酶的分离的核酸序列。

2.一种分离的多核苷酸，选自：

a)含有有效地连接至重组启动子的编码 SEQ ID NO: 2 的多肽的核苷酸序列的分离的多核苷酸；

b)含有有效地连接至重组启动子的 SEQ ID NO: 1 的分离的多核苷酸；和

c)与(a) 或(b)的多核苷酸序列互补的，有效地连接至重组启动子的分离的多核苷酸。

3.一种 DNA 构建体，包含在 5' 至 3' 方向的转录中作为有效结合组分的植物细胞中的功能性启动子、编码具有序列 SEQ ID NO: 2 的 1-脱氧-D-木酮糖 5-磷酸还原异构酶的核酸序列和转录终止序列。

4.一种宿主细胞，包含一种 DNA 构建体，该 DNA 构建体包含在 5' 至 3' 方向的转录中作为有效结合组分的植物细胞中的功能性启动子、编码具有序列 SEQ ID NO: 2 的 1-脱氧-D-木酮糖 5-磷酸还原异构酶的核酸序列和转录终止序列。

5. 根据权利要求 4 所述的宿主细胞，其中该宿主细胞是植物细胞。

6. 一种植物细胞，其含有 DNA 构建体，该 DNA 构建体包含在 5' 至 3' 方向的转录中作为有效结合组分的植物细胞中的功能性启动子、编码具有序列 SEQ ID NO: 2 的 1-脱氧-D-木酮糖 5-磷酸还原异构酶的核酸序列和转录终止序列。

7. 一种在植物中改变类异戊二烯含量的方法，包括用构建体转化

所述宿主植物，该构建体中含有作为有效连接组分的植物中的功能性转录起始区域、含有选自以下核酸序列的核酸序列以及转录终止区域：

- a) 编码 SEQ ID NO: 2 的多肽的核苷酸序列；
- b) 含有 SEQ ID NO: 1 的核苷酸序列；以及
- c) 与 (a) 或 (b) 序列互补的核酸序列。

8. 一种根据权利要求 7 所述的在植物中改变类异戊二烯化合物含量的方法，其中所述核酸序列处于有义方向。

9. 一种根据权利要求 8 所述的方法，其中的类异戊二烯化合物含量得到了提高。

10. 一种根据权利要求 7 所述的在植物中改变类异戊二烯含量的方法，其中所述核酸序列处于反义方向。

11. 一种根据权利要求 10 所述的方法，其中的类异戊二烯含量得到了减低。

12. 一种在植物细胞中生成目的类异戊二烯化合物的方法，所述方法包括获得转化的植物，所述植物在其基因组中具有并表达，该基因组包括：

一种含有选自以下序列的 DNA 序列的初级构建体：

- a) 编码 SEQ ID NO: 2 的多肽的核苷酸序列；
- b) 含有 SEQ ID NO: 1 的核苷酸序列；以及
- c) 与 (a) 或 (b) 序列互补的核酸序列,以及

至少一种含有编码参与生成某一特异性类异戊二烯的蛋白质的 DNA 序列的次级构建体，并且该次级构建体与植物细胞中的功能性转录起始区域有效地连接。

13. 一种根据权利要求 12 所述的方法，其中所述蛋白质参与类异戊二烯类化合物的生成，该类异戊二烯化合物选自生育酚、类胡萝卜素、单萜、双萜和质体醌。

14. 一种在宿主植物细胞中提高非甲羟戊酸型类异戊二烯的生物合成流的方法，所述的方法包括用构建体转化所述的宿主植物，在所使用的构建体中含有作为有效连接组分的植物中的功能性转录起始区域、含有选自以下的 DNA：

- a) 编码 SEQ ID NO: 2 的多肽的核苷酸序列；
- b) 含有 SEQ ID NO: 1 的核苷酸序列；以及
- c) 与 (a) 或 (b) 序列互补的核酸序列，
以及转录终止区域。

15. 一种在植物中调制疾病抗性的方法，包括：

培育在其基因组中含有一种构建体的植物，该构建体提供含有选自以下核苷酸的核酸序列的表达：

- a) 编码 SEQ ID NO: 2 的多肽的核苷酸序列；
- b) 含有 SEQ ID NO: 1 的核苷酸序列；以及
- c) 与 (a) 或 (b) 序列互补的核酸序列。

编码参与类异戊二烯合成的蛋白质的核酸序列

引言

本专利申请要求美国在先申请 60/129, 899 (于 1999 年 4 月 15 日提交) 以及美国在先申请 60/146, 461 (于 1999 年 7 月 30 日提交) 的优先权。

技术领域

本发明涉及核酸与氨基酸的序列和构建体, 及其相关方法。

背景

类异戊二烯类化合物广泛地存在于所有生物体中。植物合成多种系列的类异戊二烯类化合物, 超过 22, 000 种之多 (Connolly 和 Hill (1992) 类萜词典 (Dictionary of Terpenoids), Chapman 和 Hall, New York, NY)。在植物中, 类异戊二烯类化合物在特定的细胞功能诸如固醇的生成中起重要作用, 促成真核细胞膜的构造以及存在于泛醌和质体醌、生长调节剂 (如脱落酸、赤霉素、油菜素类固醇) 或光合色素叶绿素和类胡萝卜素的非环聚异戊二烯化合物的侧链。尽管其它的植物性类异戊二烯类化合物的生理作用 (比如, 多种系列的次级代谢产物的生理作用) 并不明显, 但是已知有些类异戊二烯类化合物在介导对不同的环境挑战的适应性反应中起关键作用。尽管类异戊二烯类化合物的结构和功能具有显著的多样性, 但是它们都源自单一种代谢前体, 即异戊烯二磷酸 (IPP) (Wright, (1961) 生物化学年鉴 (Annu. Rev. Biochem.), 20: 525-548; Spurgeon 和 Porter, (1981) “Biosynthesis of Isoprenoid Compounds”, Porter 和 Spurgeon 编辑, 第一卷, 第 1-46 页, John Wiley New York 出版)。

在高等植物的叶绿体中, 存在着源自类异戊二烯途径的若干独特和相互关联的生化途径, 它们导致次级代谢产物 (包括生育酚) 的生成。生育酚 (即, 维生素 E) 不仅在植物中起着至关重要的作用, 而且

从哺乳动物的营养角度来看也是重要的。在质体中，生育酚占总醌库的量达到 40%。

生育酚和生育三烯酚（不饱和生育酚的衍生物）是众所周知的抗氧化剂，在保护细胞免受自由基损害以及预防多种疾病（包括心脏病、癌症、白内障、视网膜病、老年性痴呆和神经变性）中起重要的作用，而且已经表明它们对关节炎症状以及在抗衰老中具有有益的效应。维生素 E 被用于鸡饲料中，以提高鸡肉的货架寿命、外观、香味和氧化稳定性，以及将母育酚从饲料转移到鸡蛋中。已经表明，维生素 E 对家畜动物的正常生育是必需的，并改善家畜动物的总体表现和增强其免疫力。在动物饲料中补充维生素 E 还可赋予奶制品氧化稳定性。

天然生育酚作为补充物，在过去的三年里，对其需求一直在以 10-20% 的速度稳定地增长。目前，对天然生育酚的需求超过了供给，已知天然生育酚的生物活性，高于合成得到的生育酚的外消旋混合物。天然生育酚都是 d-立体异构体，而合成 α -生育酚是八种 d, l- α -生育酚异构体的混合物，只有其中的一种（占 12.5%）与天然 d- α -生育酚完全相同。天然 d- α -生育酚，与其它天然生育酚或生育三烯酚相比，具有最高的维生素 E 活性（1.49 IU/mg）。合成 α -生育酚的维生素 E 活性为 1.1 IU/mg。1995 年，世界市场对生育酚精制原料的需求为 10.2 亿美元；其中合成物质占市场的 85-88%，剩下的 12-15% 为天然物质。天然生育酚和生育三烯酚的最佳来源物为植物油和谷类产品。当前，大部分的天然维生素 E 是产自 γ -生育酚，而 γ -生育酚又出自豆油加工，随后通过化学修饰可将其转化为 α -生育酚（ α -生育酚表现出最高的生物活性）。

不通过化学修饰而提高植物中生育酚和生育三烯酚（特别是那些能够直接使用的更希望得到的化合物）水平的方法，将有助于本领域中，这是由于这些分子表现出更好的功能性和生物可利用性。

另外，希望得到这样的方法，其可增加宿主植物细胞中其它类异戊二烯衍生化合物的产量。而且，还需要在宿主植物细胞中产生特定的类异戊二烯化合物的方法。

发明概述

本发明涉及 D-1-脱氧木酮糖 5-磷酸还原异构酶 (dxr)，更具体地，涉及 dxr 的多核苷酸和多肽。本发明的多核苷酸和多肽包括源自真核细胞来源物的那些多核苷酸和多肽。

因而，本发明的一个方面涉及编码 D-1-脱氧木酮糖 5-磷酸还原异构酶蛋白质的分离的多核苷酸序列。更具体地，本发明提供了编码来自植物源的 dxr 蛋白质的分离的核酸序列。

本发明的另一个方面涉及寡核苷酸，其中包括部分或完整的 dxr 编码序列。

本发明的又一个方面提供了可用于转录或者转录及翻译（表达）dxr 的重组 DNA 构建体。更具体地，本发明提供了能够在宿主细胞中转录或者转录及翻译的构建体。

在本发明的另一个方面中，提供了在宿主细胞或其后代中产生 dxr 的方法。更具体地，用可以用于转录或者转录及翻译 dxr 的 DNA 构建体对宿主细胞进行了转化或转染。含有 dxr 的重组细胞也是本发明的一部分。

另一方面，本发明涉及使用多核苷酸和多肽序列改变宿主细胞（尤其是宿主植物细胞）中的类异戊二烯含量的方法。在本文中构思了含有如此改变的类异戊二烯含量的植物细胞。

通过表达 dxr 而得到的被修饰的植物、种子和油也被认为是本发明的一部分。

附图简述

图 1 提供了拟南芥菜属 dxr 序列和大肠杆菌 (E. coli) dxr 序列之间的氨基酸顺序。

图 2 提供了甲羟戊酸以及非甲羟戊酸途径的类异戊二烯途径流程图。

发明详述

本发明尤其提供了用于改变（例如提高和减低）类异戊二烯水平和/或调整它们在宿主细胞中比例的组合物和方法。更具体地，本发明

提供用于调整宿主植物细胞中类异戊二烯含量的多核苷酸、多肽及其使用方法。

类异戊二烯类化合物源自 5-碳构造单元、异戊二烯二磷酸 (IPP)，它是通用的异戊二烯单元和常见的类异戊二烯前体。类异戊二烯类化合物包含一组在结构上多种多样的化合物，其可分为两类：初级和次级代谢产物 (Chappell, 1995, 植物生理学和植物分子生物学年鉴 (Annu. Rev. Plant Physiol. Plant Mol. Biol.), 46: 521-547)。初级代谢产物包括这样一些类异戊二烯类化合物，它们对膜完整性、光保护、协调发育程序以及将生化功能锚定于特定的膜系统是必需的。这样的初级代谢产物包括但不限于固醇、类胡萝卜素、叶绿素、生长调节剂，以及多萜醇、醌和蛋白质的聚戊二烯醇 (polyprenol) 取代物。次级代谢产物介导植物与环境间的重要相互作用，但对植物的生存并不是必需的。次级代谢产物包括但不限于生育酚、单萜类化合物、倍半萜类化合物和二萜类化合物。

许多年来，一直认为 IPP 是通过众所周知的乙酸/甲羟戊酸途径合成的。但是，最近的研究已证实还存在一种不依赖于甲羟戊酸的替代途径，用于 IPP 的生物合成 (Horbach 等人, (1993) FEMS Microbiol. Lett. 111:135-140; Rohmer 等人, (1993) 生物化学杂志 (Biochem. J), 259: 517-524)。该 IPP 生物合成的非甲羟戊酸途径最初是在细菌中被表征的，后来在绿藻和高等植物中也被表征 (关于最近的综述，见 Lichtenthaler 等人, (1997) 植物生理学 (Physiol. Plant.), 101: 642-652 以及 Eisenreich 等人, 1998, 化学生物学 (Chem. Biol.), 5: R221-R223)。这个新的不依赖甲羟戊酸途径的第一步反应是 (羟乙基) 硫胺素 (源自丙酮酸) 与 D-甘油醛 3-磷酸的 C 1 醛基的缩合反应，生成 D-1-脱氧木酮糖 5-磷酸 (Broers, (1994) 博士论文: Eidgenössische Technische Hochschule, 苏黎世, 瑞士; Rohmer 等人, (1996) 美国化学学会杂志 (J. Am. Chem. Soc.), 118: 2564-2566)。在大肠杆菌中，D-1-脱氧木酮糖 (最可能是以 D-1-脱氧木酮糖 5-磷酸的形式存在) 被有效地引入到甲基萜醌和泛醌的戊烯侧链中 (Broers, (1994) 同上; Rosa Putra 等人, 1998, Tetrahedron Lett., 39: 23-26)。在植物中，也报道了 D-1-脱氧木酮糖引入到类异戊二烯类化合物中 (Zeidler 等人, 1997, Z.

Naturforsch 52c: 15-23; Arigoni 等人, 1997, 美国国家科学院院刊 (Proc. Natl. Acad. Sci. USA), 94: 10600-10605; Sagner 等人, (1998) 化学通讯 (Chem. Commun.), 2: 221-222)。另外, D-1-脱氧木酮糖还被描述为硫胺素和吡哆醇的生物合成前体。在大肠杆菌中 D-1-脱氧木酮糖是硫胺素上噻唑环的连续五个碳单元 (C4'-C4-C5-C5'-C5'') 的前体分子 (Therisod 等人, 1998, 生物化学与生物物理学研究通讯 (Biochem. Biophys. Res. Comm.), 98: 374-379; David 等人 (1981) 美国化学学会杂志, 103: 7341-7342), 在高等植物叶绿体中也是如此 (Julliard 和 Douce, 1991, 美国国家科学院院刊, 88: 2042-2045)。还很好地记载了对 D-1-脱氧木酮糖在大肠杆菌中生物合成吡哆醇中的作用 (Hill 等人, (1989) 美国化学学会杂志, 111: 1916-1917; Kennedy 等人, (1995) 美国化学学会杂志, 117: 1661-1662; Hill 等人, (1996) 生物化学杂志 (J. Biol. Chem.), 271: 30426-30435)。最近, 已报道在细菌 (Sprenger 等人, (1997) 美国国家科学院院刊, 94: 12957-12962; Lois 等人, (1998) 美国国家科学院院刊, 95: 2105-2110) 和植物 (Lange 等人, (1998) 美国国家科学院院刊, 95: 2100-2104; Bouvier 等人, (1998) 植物生理学 (Plant Physiol.), 117: 1423-1431) 中已克隆了编码 1-脱氧-D-木酮糖 5-磷酸合酶的基因。图 2 提供了说明类异戊二烯途径的流程图。

尽管现在还没有表征出 1-脱氧-D-木酮糖 5-磷酸与 IPP 之间的中间体, 但是 Rohmer 与其同事已经提出将 2-C-甲基-D-赤藓醇 4-磷酸作为细菌中类异戊二烯生物合成的第一个关键前体 (Duvold 等人, (1997) Tetrahedron Lett., 38: 4769-4772; Duvold 等人, 1997, Tetrahedron Lett., 38: 6181-6184)。催化 1- D-脱氧-D-木酮糖 5-磷酸转变为 2-C-甲基-D-赤藓醇 4-磷酸的 1-脱氧-D-木酮糖 5-磷酸还原异构酶, 最近已在大肠杆菌中被克隆和表征 (Takahashi 等人, 1998, 美国国家科学院院刊, 95: 99879-99884)。最近 Sagner 等人分别在 Tetrahedron Lett., (1998), 39: 23-26 和化学通讯, (1998), 2: 221-222 中报道了通过 1-脱氧-D-木酮糖 5-磷酸的分子内重排在植物中生物合成 2-C-甲基-D-赤藓醇。

本发明提供了参与从 1-脱氧木酮糖-5-磷酸生成 2-C-甲基-D-赤藓醇-4-磷酸的多核苷酸和多肽的序列, 其被称为 1-脱氧-D-木酮糖 5-磷酸

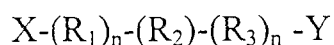
还原异构酶或 dxr。本发明还提供了在宿主细胞中产生表达改变了的 dxr 的构建体和方法，以及修饰类异戊二烯途径（包括对通过类异戊二烯途径的生物合成流的修饰）的方法，和在宿主细胞中产生特定种类 of 类异戊二烯类化合物的方法。

分离的多核苷酸、蛋白质和多肽

本发明的第一个方面涉及分离的 dxr 多核苷酸。本发明的多核苷酸序列包括编码本发明多肽（它具有选自序列表中列出的序列组的推定氨基酸序列）的分离的多核苷酸，还涉及与这些序列及其变异体紧密相关的其它多核苷酸序列。

本发明提供了一种多核苷酸序列，其整个序列与在序列表中列出的每个编码序列完全相同。本发明还提供了成熟多肽或其片段的编码序列，以及与其它编码序列诸如那些编码导肽或分泌序列、前-、原-或前原-蛋白质序列处于同一读框中的成熟多肽或其片段的编码序列。该多核苷酸还可包含非编码序列，例如包括但不限于非编码的 5' 和 3' 序列，诸如转录但不翻译的序列、终止信号、核糖体结合位点、稳定 mRNA 的序列、内含子、多腺苷酸化信号和编码所添加氨基酸的添加编码序列。例如，可以将一个标记序列包括在内，以促进融合多肽的纯化。本发明的多核苷酸还包括含有结构基因和控制基因表达的天然结合序列的多核苷酸。

本发明还包括下式的多核苷酸：



其中，位于 5' 末端的 X 是氢，位于 3' 末端的 Y 是氢或金属， R_1 和 R_3 为任何核酸残基，n 是一个介于 1-3000 之间的整数（优选介于 1-1000 之间）， R_2 是本发明的核酸序列，尤其是选自序列表列出的核酸序列并优选 SEQ ID NO: 1 中的那些核酸序列。在上式中， R_2 是有方向性的，其 5' 末端残基位于左边，结合于 R_1 ，而其 3' 末端残基位于右边，结合于 R_3 。用一组 R（R 大于 1）表示的任何核酸残基片段，可以是杂聚物，也可以是均聚物，优选为杂聚物。

本发明还涉及本文所述的多核苷酸的变异体，它们编码本发明的多肽的变异体。那些本发明的多核苷酸片段的变异体，可被用于合成

本发明的全长多核苷酸。优选的实施方案为编码多肽变异体的多核苷酸，其中对本发明的多肽序列的 5-10，1-5，1-3，2，1 或无氨基酸残基被以任何组合替代、添加或缺失。特别优选沉默的替代、添加和缺失，因为它们不会改变该多核苷酸或多肽的性能或活性。

本发明的进一步优选的实施方案为：在多核苷酸全长内与编码本发明多肽的多核苷酸有至少 50%、60%或 70%相同性的多核苷酸，以及与这种多核苷酸互补的多核苷酸。更优选含有其全长与编码本发明多肽的多核苷酸及其互补多核苷酸有至少 80%相同性的区域的多核苷酸。在这一点上，尤其优选其全长与编码本发明多肽的多核苷酸及其互补多核苷酸有至少 90%相同性，特别优选那些有至少 95%相同性的多核苷酸序列。进一步地，非常优选那些有至少 97%相同性的多核苷酸序列，而且尤其非常优选那些有至少 98%和 99%相同性的多核苷酸序列，其中最优选那些有至少 99%相同性的多核苷酸序列。

优选的实施方案为这样的多核苷酸，它们所编码的多肽基本上保留了与成熟多肽（由序列表中列出的多核苷酸编码）相同的生物学功能或活性。

本发明进一步涉及可与上述序列杂交的多核苷酸。更具体地，本发明涉及在严格条件下可与上述多核苷酸杂交的多核苷酸。本文所用的术语“严格条件”和“严格杂交条件”是指在序列间具有至少 95%（优选至少 97%）相同性的情况下，一般会发生杂交反应。严格杂交条件的一个例子是，在含有 50%甲酰胺、5×SSC（150mM NaCl, 15mM 柠檬酸三钠）、50mM 磷酸钠（pH7.6）、5×Denhardt's 溶液、10%硫酸葡聚糖和 20 毫克/毫升变性剪切鲑精 DNA 的溶液中，在 42℃温育过夜，随后在 0.1×SSC 中于大约 65℃冲洗杂交载体。其它的杂交和冲洗条件是公知的，并且在 Sambrook 等人的分子克隆：实验室手册（Molecular Cloning: A Laboratory Manual），Cold Spring Harbor, NY（1989）第二版，尤其是第 11 章中有例子。

本发明还提供一种基本上由一种多核苷酸序列组成的多核苷酸，这样的多核苷酸可如此获得：在严格杂交条件下，用在多核苷酸序列表中列出的多核苷酸序列或其片段序列作为探针，筛选含有该种多核苷酸序列的完整基因的合适的文库；然后分离所述的核苷酸序列。用

于获得这样一种多核苷酸的片段包括例如本文中所述的探针和引物。

正如本文所讨论的关于本发明的多核苷酸测定法,例如,本发明的多核苷酸可被用作 RNA、cDNA 或基因组 DNA 的杂交探针,以分离全长 cDNA 或编码多肽的基因组克隆,以及分离与序列表中列出的多核苷酸具有高度序列相似性的其它基因的 cDNA 或基因组克隆。这样的探针一般含有至少 15 个碱基。优选地,这样的探针具有至少 30 个碱基并可以具有至少 50 个碱基。尤其优选的探针将具有 30-50 个碱基,其中包括 30 和 50 个在内。

含有或含于序列表中列出的多核苷酸序列的每一种基因,其编码区可以利用序列表中提供的 DNA 序列筛选,以合成寡核苷酸。然后,使用与本发明的基因的序列具有序列互补性的标记寡核苷酸,对 cDNA、基因组 DNA 或 mRNA 的文库进行筛选,以鉴定可与探针杂交的文库中的成员。例如,制备与 dxr 序列相应的合成的寡核苷酸。以寡核苷酸为引物,以聚合酶链式反应(PCR)技术获得 dxr 基因的 5' 和 3' 末端序列。或者,在可以根据特定的 dxr 肽制备出低简并性寡核苷酸的情况下,也可以将这样的探针直接用于筛选基因文库,以获得 dxr 基因序列。尤其是,筛选噬菌体载体中的 cDNA 文库有助于这种方法,因为杂交的背景较低。

一般地,使用核酸探针所获得的 dxr 序列显示出,靶 dxr 序列与用作探针的编码序列之间 60-70% 的序列相同。但是,也可以获得低至 50-60% 序列相同性的长序列。核酸探针可以是长的核酸序列片段,也可以是较短的寡核苷酸探针。当使用较长的核酸片段作为探针时(大于大约 100 bp),为了从靶样品中获得与探针序列具有 20-50% 偏差(即,50-80% 的序列同源性)的序列,在进行筛选时可以将杂交的严格性降低。寡核苷酸探针可以比编码 dxr 酶的完整核酸序列短许多,但至少也应该为大约 10 个,优选至少为大约 15 个,更优选至少为大约 20 个核苷酸。当使用较短的区域作探针时,与较长的区域相比,希望能达到更高的序列相同性程度。因此,希望鉴定出高度保守的氨基酸序列区,以设计寡核苷酸探针,用于探测和回收其它相关的 dxr 基因。较短的探针对于聚合酶链式反应(PCR)常常特别有用,并且在可以鉴定出高度保守序列的情况下尤其如此(参见, Gould 等人, 1989, 美国国家科

学院院刊, 86: 1934-1938)。

本发明的另一方面涉及 dxr 多肽。这种多肽包括: 在序列表中列出的分离的多肽及其多肽和片段, 尤其是那些显示 dxr 活性并且与从序列表所列序列中选出的多肽序列具有至少 50%、60% 或 70%, 优选至少 80%, 更优选至少 90%, 最优选至少 95% 相同性的多肽, 还包括这些多肽的组成部分, 其中多肽的组成部分优选包含至少 30 个氨基酸, 更优选包含至少 50 个氨基酸。

“相同性”, 正如本领域所熟知的, 是两种或更多种多肽序列间或者两种或更多种多核苷酸序列间的相互关系, 其可通过比较这些序列而确定。在本领域中, “相同性” 也指多肽或多核苷酸序列间的序列相关性程度, 其可通过对这些序列一串串地进行比对而确定。“相同性” 可以容易地通过已知的方法计算出, 这些方法包括但不限于: 在计算分子生物学 (Computational Molecular Biology, Lesk, A. M. 编辑, 牛津大学出版社, 纽约, 1988); 生物计算: 信息学和基因组计划 (Biocomputing: Informatics and Genome Projects, Smith, D. W. 编辑, Academic Press, 纽约, 1993); 序列数据的计算机分析 (第一部分) (Computer Analysis of Sequence Data, Part I, Griffin, A. M. 和 Griffin, H. G. 编辑, Humana Press, New Jersey, 1994); 分子生物学中的序列分析 (Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987); Sequence Analysis Primers, Gribscov, M. 和 Devereux, J. 编辑, Stockton Press, 纽约, (1991); 以及 Carillo, H. 和 Lipman, D., SIAM 应用计算杂志 (SIAM J Applied Math.), 48: 1073 (1988) 中描述的那些方法。在设计确定相同性的方法时, 目的就是给出待测试序列间的最大匹配。而且, 确定相同性的方法被编写在可公开获得的程序中。可用于确定两种序列间相同性的电脑程序包括但不限于: GCG (Devereux, J. 等人, 1984, 核酸研究 (Nucleic Acids Research), 12 (1): 387); 一组五个 BLAST 的程序, 其中的三个用于核酸序列查询 (BLASTN、BLASTX 和 TBLASTX), 两个用于蛋白质序列查询 (BLASTP 和 TBLASTN) (Coulson, 1994, 生物技术趋势 (Trends in Biotechnolgy), 12: 76-80; Birren 等人, 基因组分析 (Genome Analysis), 1: 543-559 (1997))。BLAST X 程序可以从 NCBI 和其它途径公开获

得 (BLAST 手册, Altschul, S.等人, NCBI NLM NIH, Bethesda, MD 20894; Altschul, S.等人, 分子生物学杂志 (J. Mol. Biol), 215: 403-410 (1990))。众所周知的 Smith Waterman 算法也可用于确定相同性。

多肽序列比较参数一般包括:

算法: Needleman 和 Wunsch, 分子生物学杂志, 48: 443-453 (1970)

比较矩阵: BLOSSUM62, 来自 Hentikoff 和 Hentikoff, 美国国家科学院院刊, 89: 10915-10919 (1992)

空位处罚分: 12

空位长度处罚分: 4

能够采用这些参数的一个程序, 可从 Genetics Computer Group, Madison, Wisconsin 公开获得, 这个程序被称为“空位”程序。上述参数, 连同对末端空位没有处罚分, 就是进行肽比较的缺省设置参数。

多核苷酸序列比较的参数包括:

算法: Needleman 和 Wunsch, 分子生物学杂志, 48: 443-453 (1970)

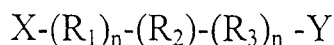
比较矩阵: 匹配 = +10, 错配 = 0

空位处罚分: 50

空位长度处罚分: 3

能够采用这些参数的一个程序, 可从 Genetics Computer Group, Madison, Wisconsin 公开获得, 这个程序被称为“空位”程序。上述参数就是进行核酸比较的缺省设置参数。

本发明还包括下式的多肽:



其中, 位于氨基末端的 X 是氢, 位于羧基末端的 Y 是氢或金属, R_1 和 R_3 为任何氨基酸残基, n 是一个介于 1-1000 之间的整数, R_2 是本发明的氨基酸序列, 尤其是选自序列表中列出的氨基酸序列, 优选在 SEQ ID NO: 2 中提供的序列编码的那些氨基酸序列。在上式中, R_2 是有方向性的, 其氨基末端残基位于左边, 结合于 R_1 , 而其羧基末端残基位于右边, 结合于 R_3 。用一组 R (R 大于 1) 表示的任何氨基酸残基片段, 可以是杂聚物, 也可以是均聚物, 优选为杂聚物。

本发明的多肽包括由多核苷酸编码的分离的多肽, 其中的多核苷酸含有选自本文序列表中列出的序列。

本发明的多肽可以是成熟的蛋白质或者可以是融合蛋白质的一部分。

这些多肽的片段和变异体也被认为是本发明的一部分。一个片段就是一个变异体多肽，它所具有的氨基酸序列与前面描述的多肽的氨基酸序列的一部分而不是全部完全相同。这些片段可以独立存在，也可以含于一个更大的多肽中，其中该片段形成一部分或一个区域，最优选是形成单一连续区域。优选的片段为具有生物学活性的片段，它们可以介导本发明多肽的活性，其中包括那些具有相似活性或提高活性或降低活性的片段。本发明还包括那些在动物尤其是人中具有抗原性或免疫原性的片段。

这些多肽的变异体还包括与序列表中列出的序列不同的多肽，这些不同是由保守氨基酸的取代产生的，即一个氨基酸残基被另一个具有相似性质的氨基酸残基取代。一般地，这中取代发生在 Ala、Val、Leu 和 Ile 之间，Ser 和 Thr 之间，Asp 和 Glu 之间，Asn 和 Gln 之间，Lys 和 Arg 之间，Phe 和 Tyr 之间。尤其优选其中 5-10、1-5、1-3 个或者 1 个氨基酸被以任何组合取代、缺失或添加的变异体。

那些是本发明多肽片段的变异体，通过肽合成可用于产生相应的全长多肽。因而，可将这些变异体用作生成本发明的全长多肽的中间体。

本发明的多核苷酸和多肽可用于，例如，转化宿主细胞诸如植物宿主细胞，正如本文所进一步讨论的。

本发明还提供了编码多肽的多核苷酸，其中的多肽为成熟蛋白质再加上添加的氨基或羧基末端氨基酸，或者成熟的多肽中的氨基酸（例如，当蛋白质的成熟形式具有多于一个多肽链时）。这种序列可以，例如，在蛋白质从前体到成熟形式的转变加工中起作用，帮助蛋白质转运，缩短或延长蛋白质半寿期，或者在蛋白质的测定或生产中简化操作。预期可使用胞内酶从成熟的蛋白质中除去任何添加的氨基酸。

具有融合到一种或多种前序列（prosequence）的成熟形式的多肽的前体蛋白质，可能是该多肽的无活性形式。一般地，当前序列被除去时，无活性前体被激活。在激活之前可以将一部分或者全部前序列除去。这种前体蛋白质一般称为前蛋白质。

所使用的构建体和方法

我们尤其感兴趣的是将核苷酸序列用于重组 DNA 构建体，以指导本发明的 dxr 序列在宿主细胞中的转录或转录及翻译（表达）。表达型构建体一般含有宿主细胞中的功能性启动子，该启动子有效地连接到本发明的编码 dxr 的核酸序列，还含有宿主细胞中的功能性转录终止区。在本发明中尤其感兴趣的宿主细胞包括但不限于真菌细胞、酵母细胞、细菌细胞、哺乳动物细胞和植物细胞。

当序列的排列方式使得第一种核酸序列影响到第二种核酸序列的功能时，这两种核酸序列就是“有效地连接”或“有效地结合”。优选地，这两种序列是单一连续核酸分子的一部分，更优选地，这两种序列相邻。例如，如果一个启动子调节或介导一个基因在细胞中的转录，那么，所述启动子就与所述基因进行了有效地连接。

本领域的技术人员会认识到，在植物细胞中有许多功能性启动子，并且已在文献中对它们进行了描述。还展望了叶绿体和质体特异性启动子，叶绿体或质体中功能性启动子，以及在叶绿体或质体中有效的启动子。

一组植物功能性启动子为组成型启动子如 CaMV35S 或 FMV35S 启动子，它们在大多数植物器官中产生高水平的表达。已将 CaMV35S 和 FMV35S 启动子的增强或复制版本用于本发明的实践中（Odell 等人，1985，自然（Nature），313：810-812；Rogers，美国专利 5378619）。另外，可优选让 dxr 基因在植物的特定组织（如叶、茎、根、块茎、种子、果实等）中表达，并且所选择的启动子应具有所需要的组织和发育特异性。

尤其感兴趣的是，使本发明的核酸序列从转录起始区开始表达，并优先在植物种子组织中表达。这种优先在种子中转录起始的序列例子包括那些源自编码植物储藏蛋白质基因序列或者油料种子中参与脂肪酸生物合成的基因的序列。这种启动子的例子包括来自诸如 napin（Kirdle 等人，1991，种子科学研究（Seed Sci. Res.），1：209-219）、菜豆蛋白、玉米醇溶蛋白、大豆胰蛋白酶抑制剂、ACP、硬脂酰-ACP 去饱和酶、大豆 β -conglycin 的 α' 亚单位（大豆 7s，（Chen 等人，1986，

美国国家科学院院刊, 83: 8560-8564)) 和油质蛋白的基因的 5' 调节区。

指导具有 dxr 的蛋白质定位于特定的亚细胞隔离空间例如线粒体、内质网、液泡、叶绿体或其它质体性隔离空间中, 也许是有益的。例如, 当将本发明的目的基因靶向质体诸如叶绿体以表达时, 构建体也要使用指导基因至质体的序列。这种序列在本文中称为叶绿体转运肽 (CTP) 或质体转运肽 (PTP)。这样, 在没有将目的基因直接插入到质体中的情况下, 表达型构建体将另外含有编码转运肽的基因, 从而指导目的基因至质体。叶绿体转运肽可以是源自目的基因, 也可以是源自具有 CTP 的异种序列。这种转运肽在本领域中是公知的。参见例如, Von Heijne 等人, 植物分子生物学研究 (Plant Mol. Biol. Rep.), 9: 104-126 (1991); Clark 等人, 生物化学杂志 (J. Biol. Chem.), 264: 17544-17550 (1989); della-Cioppa 等人, 植物生理学 (Plant Physiol.), 84: 965-968 (1987); Romer 等人, 生物化学与生物物理学研究通讯 (Biochem. Biophys. Res. Commun.), 196: 1414-1421 (1993); 以及 Shah 等人, 科学 (Science), 233: 478-481 (1986)。

根据目的用途, 构建体可以含有编码整个 dxr 蛋白质或其一个组成部分的核酸序列。例如, 当需要反义抑制给定的 dxr 蛋白质时, 就不需要完整的 dxr 序列。而且, 当用于构建体中的 dxr 序列是旨在用作探针时, 则制备只含有 dxr 编码序列的特定组成部分 (例如, 被发现是编码高度保守的 dxr 区的序列) 的构建体可能是有益的。

技术人员会认识到, 用于抑制内源序列在宿主细胞中表达的方法有许多。这种方法包括但不限于反义抑制 (Smith 等人, 1988, 自然, 334: 724-726)、共抑制 (Napoli 等人, 植物细胞 (Plant Cell), 2: 279-289)、核酶 (PCT 公开说明书, WO 97/10328) 以及有义与反义的组合 (Waterhouse 等人, 1998, 美国国家科学院院刊, 95: 13959-13964)。在宿主细胞中抑制内源序列的方法一般使用了转录或转录及翻译至少一部分待抑制序列。这种序列可以与内源序列的编码区同源, 也可以与内源序列的非编码区同源。

本发明的植物表达构建体也可以提供调节转录本的终止区。转录本的终止区可以由编码 dxr 的 DNA 序列提供, 或者由不同基因来源的

方便的转录终止区提供,例如,与转录本起始区天然结合的转录本终止区。技术人员会认识到,任何能够在植物细胞中终止转录的方便的转录本终止区,都可以在本发明的构建体中予以采用。

或者,也可以将构建体制备为可以指导 dxr 序列直接从宿主植物细胞的质体表达。这种构建体和方法为本领域所公知,并一般描述于,例如, Svab 等人, 1990, 美国国家科学院院刊, 87: 8526-8530; Svab 和 Maliga, 1993, 美国国家科学院院刊, 90: 913-917; 以及美国专利 5693507 中。

本发明的构建体也可用于这样的方法: 用表达添加基因(参与类异戊二烯化合物生成)的添加构建体, 通过类异戊二烯途径改变代谢流。这样的序列包括但不限于 1-脱氧木酮糖 5-磷酸合酶。

而且, 本发明的构建体也能够用于转化方法中, 其中使用在生成特定的类异戊二烯化合物(诸如生育酚、类胡萝卜素、固醇、单萜、倍半萜和双萜)中提供编码蛋白质的添加核酸序列的表达的添加的构建体。参与生成类胡萝卜素的核酸序列和方法在例如描述于 PCT 公开说明书 WO 99/07867 中。参与生成生育酚的核酸序列包括但不限于 γ -生育酚甲基转移酶(Shintani 等人, 1998, 科学, 282(5396): 2098-2100)、生育酚环化酶、生育酚甲基转移酶、叶绿基异戊二烯基转移酶、牻牛儿苗牻牛儿苗焦磷酸氢化酶、牻牛儿苗牻牛儿苗焦磷酸合酶。

当将含有表达型构建体的重组 DNA 构建体引入到植物的细胞、组织、器官中或植物中时, 我们就认为植物的细胞、组织、器官或植物已被转化、转染或转基因。转基因或转化的细胞或植物也包括细胞或植物的后代, 以及由育种项目而产生的后代, 该育种项目在杂交试验中采用这样的转基因植物作为亲本, 并且这些后代由于 dxr 核酸序列的存在而表现出变化的表型。

具有 dxr 编码序列并作为提高或减低其表达的目的 DNA 序列的植物表达或转录型构建体, 可用于广泛多种植物生命体中。用于本发明方法的尤其优选的植物包括但不限于: 金合欢、苜蓿、茴香树、苹果树、apricot、洋葱、arugula、石刁柏、鳄梨树、芭蕉属植物、大麦、豆科植物、甜菜、黑莓、乌饭树、嫩茎花椰菜、brussels sprout、卷心菜、canola、罗马甜瓜、胡萝卜、木薯、花椰菜、芹菜、樱桃树、菊苣属植

物、cilantro、柠檬、clementines、咖啡树、玉米、棉花、黄瓜、Douglas 冷杉、茄子、苣荬菜、escarole、桉树属植物、茴香、无花果属植物、大蒜、葫芦属植物、葡萄、葡萄柚、加甜味烟草、jicama、kiwifruit、莴苣、韭葱、柠檬树、lime、Loblolly 松树、芒果、甜瓜、菌类植物、油桃、坚果、燕麦属植物、油棕榈、菜籽油油菜、okra、洋葱、柑橘属植物、一种装饰性植物、番木瓜树、欧芹、豌豆、桃树、花生、梨树、胡椒属植物、柿树、松树、菠萝、车前草、李属植物、石榴树、杨属植物、马铃薯、南瓜、quince、放射状松树、radicchio、小萝卜、木莓、水稻、黑麦、高粱、Southern 松树、大豆、菠菜、草莓、糖甜菜、甘蔗、向日葵、白薯、sweetgum、红橘树、茶树、烟草、西红柿、triticale、草皮、芜菁、一种葡萄树、西瓜、小麦、薯蓣属植物和 zucchini。

尤其优选这样的植物，它们参与生成食用或工业用途的植物油。最特别优选温带油料作物。感兴趣的温带油料作物包括但不限于：油菜籽（Canola 和高芥子酸变种）、向日葵、红花、棉花、大豆、花生、椰子和油棕以及玉米。依据引入重组构建体至宿主细胞内所使用的方法，也许需要其它 DNA 序列。重要的是，本发明适用于双子叶种类的植物和单子叶种类的植物，并可以容易地适用于新的和/或改进的转化和调节技术。

尤其感兴趣的是，在植物中使用了 dxr 构建体，以产生这样的植物或植物体部分（其中包括但不限于叶、茎、根、生殖性部分和种子），在具有转化植物细胞的植物体部分中，生育酚的含量被改变。

为了进行免疫学筛选，用纯化的蛋白质或其组成部分注射兔或小鼠就可以制备针对该蛋白质的抗体，这种制备抗体的方法为本领域人员所公知。可以制备产生单克隆抗体或多克隆抗体，尽管在一般情况下多克隆抗体对于基因分离更有用。可以进行 Western 分析，以确定在所希望的植物种类的粗提取物中有相关蛋白质的存在，例如采用针对被编码蛋白质的抗体通过交叉反应来确定。当观察到交叉反应时，通过筛选代表所希望植物种类的表达文库，将编码相关蛋白质的基因分离。表达文库可构建于多种可商购载体中，包括 λ gt11，正如在 Sambrook 等人的分子克隆：实验室手册（Molecular Cloning: A Laboratory Manual），Cold Spring Harbor Laboratory, Cold Spring Harbor,

New York 第二版（1989）中所描述的。

为了证实被鉴定为 dxr 酶的核酸序列所编码的蛋白质的活性和特异性，采用杆状病毒表达系统在昆虫细胞培养物中进行了体外测定。这种杆状病毒表达系统为本领域所公知，并且在 Lee 等人的美国专利 5348886 中有描述，本文引用其全部作为参考。

另外，可以制备其它表达型构建体，以使用不同的表达系统测定蛋白质活性。将这种表达型构建体转化到酵母或原核细胞宿主内并测定 dxr 活性。这种表达系统为本领域所公知，并可容易地通过商业途径获得。

除了在本发明中所描述的序列外，用于本发明的 DNA 编码序列也可源自藻、真菌、细菌、哺乳动物来源物、植物等。可以使用与 dxr 的保守核苷酸或氨基酸序列相应的特征序列，在已有的数据库中进行同源性搜索，从而从其它来源物诸如植物和微生物中分离相当、相关基因。也可以在 EST 数据库中进行搜索。而且，本发明还包含在功能上、酶本身方面与本文所公开的相当的 DNA 序列编码酶的用途，其中该 DNA 序列是本文中所公开的与遗传密码的简并性相一致的核酸序列的简并性相当物。使用任何这些方法所鉴定的编码序列，对其功能的证实，可以通过缺乏特异性生化反应或者已经发生突变的合适生物体（例如，聚胞蓝细菌属（*Synechocystis*）、*Schewanella*、酵母、假单胞菌属（*Pseudomonas*）、红细菌科（*Rhodobacteria*）等）的突变体的互补来实现。DNA 编码区的序列可以依据密码子的使用选择性，通过基因合成予以最优化，从而在特定的宿主中实现表达的最大化。

在获得这种转基因植物中，转化方法对于本发明并不是非常重要，并且当前有多种植物转化方法可供使用。另外，如果有更新可用于转化作物的方法，此后可以直接应用这些新方法。例如，天然对农杆菌属（*Agrobacterium*）感染敏感的许多植物种类，可以通过农杆菌属介导转化的三分或二分载体方法予以成功地转化。在许多情况下，希望使构建体的一侧或两侧都与 T-DNA 相连，优选使左侧和右侧与 T-DNA 相连，更优选使右侧相连。这在构建体使用根癌农杆菌（*A. tumefaciens*）或毛根农杆菌（*A. rhizogenes*）作为转化方式时显得尤其有用，尽管 T-DNA 边缘区在其它转化方式中也有用途。另外，已经开发了显微注

射、DNA 颗粒轰击和电穿孔技术，这些技术可以转化多种单子叶和双子叶植物种类。

在正常情况下，在 DNA 构建体中将包含这种结构基因，它具有在宿主中表达所必需的调节区并提供对转化细胞的选择。该基因可以提供对一种细胞毒试剂例如抗生素、重金属、毒素等的抗性，通过互补对营养缺陷型宿主提供原养型，提供病毒免疫力等等。根据不同宿主种类的数目，引入表达型构建体或其组分，可以使用一种或多种标记，从而对不同的宿主种类使用不同的选择条件。

在使用农杆菌属转化植物细胞时，可以使用这样一种载体，其被引入农杆菌属宿主后，可与存在于农杆菌属宿主中的 T-DNA 或 Ti-或 Ri-质粒进行同源重组。含有 T-DNA（用于重组）的 Ti-或 Ri-质粒可以被接臂（能引起瘿的形成）的，也可以被断臂（不能引起瘿的形成）的，在后一种情况下，只要 *vir* 基因存在于转化的农杆菌属宿主中就是可行的。接臂的质粒能够产生正常植物细胞与瘿的混合物。

在将农杆菌属用作转化宿主植物细胞的媒介物的情况下，将通过 T-DNA 边缘区相连的表达或转录构建体插入到一种宽宿主范围的载体中，这种载体具有在大肠杆菌以及农杆菌属中复制的能力，在文献中有对多种宽宿主范围载体的描述。通常使用的是 pRK2 或其衍生物。参见例如，Ditta 等人，1980，美国国家科学院院刊，77：7347-7351 以及 EPA 0120515，此处引用作为参考。或者，也可以将准备在植物细胞中表达的序列插入到含有独立的复制序列的载体中，其中的一个复制序列在大肠杆菌中稳定载体，另一个在农杆菌属稳定载体。见例如，MacBride 等人，1990，植物分子生物学 (Plant Mol. Bol.)，14：269-276，其中使用了 pRiHRI 复制起点 (Jouanin 等人，1985，分子普通遗传学 (Mol. Gen. Genet.)，201：370-374)，并使植物表达载体在宿主农杆菌属细胞中的稳定性增加。

包括表达型构建体和 T-DNA 的可以是一种或多种标记，它们允许对转化的农杆菌属以及转化的植物细胞进行选择。已经开发了若干种用于植物细胞的标记，诸如对氯霉素、卡那霉素、氨基糖苷 G418、潮霉素等的抗性，本发明并不要求必须使用某种标记，而是根据特定的宿主以及构建的方式优选一种或另一种标记。

在使用农杆菌属转化植物细胞时，可以将外植体与转化的农杆菌属合并在一起并温育足够的时间以使转化发生，细菌被杀死，然后将植物细胞培养于合适的选择培性培养基中。一旦愈伤组织形成，通过使用合适的植物激素（采用公知的方法）可以催促茎干形成，然后将茎干转移到生根培养基中以再生植物。接着可以让植物生长结种，然后将种子用于建立重复生产以及用于分离植物油。

有几个可能的方法，用来获得含有多个表达型构建体的本发明的植物细胞。本发明包含了任何用于生成植物的方法，该植物具有一种构建体，其具有编码本发明的表达型构建体的 DNA 序列，以及至少一种其它构建体，其具有另一种编码酶的 DNA 序列。例如，可以通过将两种表达型构建体合在单一种转化载体中，或使用独立的载体将表达型构建体与第二种构建体同时用于转化植物，其中的每一种载体均表达所需要的基因。还可以将第二种构建体引入到已经用 dxr 表达型构建体转化的植物中，或者也可以将一个是表达 dxr 构建体、另一个是表达第二种构建体的被转化植物，互相进行杂交，从而使两种构建体合于同一植物中。

可以将本发明的核酸序列用于构建体中，以提供在多种宿主细胞（真核细胞和原核细胞）中该序列的表达。本发明的宿主细胞优选包括单子叶和双子叶植物细胞。

一般地，技术人员对那些描述大分子（例如，DNA 分子、质粒等）的构建、操作和分离，产生重组生物体以及筛选和分离克隆（参见例如，Sambrook 等人的分子克隆：实验室手册（Molecular Cloning: A Laboratory Manual），Cold Spring Harbor Press（1989）；Maliga 等人，植物分子生物学方法（Methods in Plant Molecular Biology），Cold Spring Harbor Press（1985），其全部内容在此处引入作为参考；Birren 等人，基因组分析：分析 DNA（Genome Analysis: Analysing DNA），1，New York，其全部内容在此处引入作为参考）的特定条件及方法的标准资料都很熟悉。

在昆虫细胞中表达序列的方法在本领域中是公知的。杆状病毒表达载体是重组的昆虫病毒，其中已将所选择的外源基因的编码序列插入到杆状病毒启动子的后面，以取代病毒基因，例如，多角体蛋白基

因 (Smith 和 Summers, 美国专利 4745051, 其全部内容在此处引入作为参考)。杆状病毒表达载体为本领域所公知, 并在例如 Doerfler, 1968, *Curr. Top. Microbiol. Immunol.*, 131: 51-68; Luckrow 和 Summers, 1988a, 生物技术 (Bio/Technology), 6: 47-55; Miller, 1988, 微生物学年鉴 (Annual Review of Microbiology), 42: 177-199; Summers, 1988, *Curr. Comm. Molecular Biology*, Cold Spring Harbor Press, Cold Spring Harbor, N. Y.; Summers 和 Smith, 1988, 杆状病毒载体和昆虫细胞培养的方法手册 (A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures), Texas Ag. Exper. Station Bullentin No. 1555 中对其进行描述, 其全部内容在此处引入作为参考。

在真菌宿主细胞中表达目的核酸序列的方法为本领域所公知。真菌宿主细胞可以是例如, 酵母细胞或丝状真菌细胞。在酵母细胞中表达目的 DNA 序列的方法一般描述于“酵母的遗传学和分子生物学指南”(Guide to Yeast Genetics and Molecular Biolgy), Guthrie 和 Fink 编辑, 酶学方法 (Methods in Enzymology), Academic Press, Inc., 第 194 卷 (1991 年); 以及“基因表达技术”(Gene Expression Technology), Goeddel 编辑, 酶学方法, Academic Press, Inc., 第 185 卷 (1991 年)。

可用作表达宿主的哺乳动物细胞是本领域所公知的, 其中包括许多无限增殖化细胞系, 可从美国典型培养物保藏中心 (ATCC, Manassas, VA) 得到它们, 诸如 Hela 细胞、中国仓鼠卵巢 (CHO) 细胞、幼小仓鼠肾 (BHK) 细胞和若干其它细胞系。用于哺乳动物细胞的合适的启动子也是本领域所公知的, 其中包括但不限于病毒性启动子诸如来自猴病毒 40 (SV40) (Fiers 等人, 1978, 自然, 273: 113, 此处引入其全部内容作为参考)、劳斯肉瘤病毒 (RSV)、腺病毒 (ADV) 和牛乳头瘤病毒 (BPV) 的启动子。哺乳动物细胞也需要终止子序列和 poly-A 添加序列。也可以将提高表达的增强子序列包括在内, 并且也希望含有可促进基因扩增的序列 (例如, 甲氨蝶呤抗性基因)。

适于在哺乳动物细胞中复制的载体是本领域所公知的, 其中可包括病毒性复制子, 或者能够确保将编码表位的适当序列整合到宿主基因组中的序列。已经对那些极大地促进重组病毒构建的质粒载体进行了描述 (参见例如, Mackett 等人, 1984, 病毒学杂志 (J. Virol.), 49:

857; Chakrabarti 等人, 1985, 分子细胞生物学 (Mol. Cell. Biol.), 5: 3403; Moss, 哺乳动物细胞基因转移载体 (Gene Transfer Vectors for Mammalian Cells), Miller 和 Calos 编辑, Cold Spring Harbor Laboratory, N.Y., 1987 年, 第 10 页; 此处引入上述全部文献的全部内容作为参考)。

到目前为止, 我们已对本发明进行了综述, 通过参考下面的实施例, 将更容易了解本发明, 所给出的这些实施例, 仅用于说明, 而不是构成对本发明的限制。

实施例

实施例 1: 2-C-甲基-D-赤藓醇的合成

根据 Duvold 等人的 Tetrahedron Lett., 38: 4769-4772 (1997) 以及 Duvold 等人的 Tetrahedron Lett., 38: 6181-6184 (1997), 并进行相应的改变以适应更大量生产, 合成了具有大约 80% (如有错误, 可予更正) 的 2-C-甲基-D-赤藓醇。将 3-甲基-2(5H)-呋喃糖 (200mg, 2mmol) 的无水乙醚 (20ml) 溶液, 在氩气下, 于 0°C 在 15 分钟长的时间, 加入到搅拌的 LiAlH_4 (46mg, 1.2mmol) 无水乙醚 (20ml) 悬浮液中。将反应混合物在 0°C 下再搅拌 2 小时。将饱和 NH_4Cl 溶液 (2ml) 缓慢地加入到其中, 直到将过量 LiAlH_4 破坏。在用 1M HCl 溶液酸化直到所有的铝盐被溶解后, 用乙酸乙酯 (6×12ml) 对液相进行萃取。将合并的有机相层用饱和盐水冲洗, 然后用无水 Na_2SO_4 干燥。在减压条件下将溶剂除去后, 将此二醇粗品 (177 mg) 溶于二氯甲烷 (20 ml), 然后在催化量的二甲胺吡啶 (12 mg) 存在下在乙酸酐/三乙胺混合液 (2: 3, 体积/体积, 1ml) 中直接乙酰化 15 分钟。在减压条件下将溶剂和过量的试剂蒸发掉。采用急骤柱层析 (Flash column chromatography) (Still 等人, 1978) (己烷/乙酸乙酯, 4: 1, 体积/体积), 得到纯的双乙酸酯 (330 mg, 86%)。于 0°C 搅拌下, 在有手性 osmylation 试剂 AD-混合物-b (2.5g) 和 $\text{CH}_3\text{SO}_2\text{NH}_2$ (152mg, 1.6mmol) 存在下在叔丁醇/水 (1: 1, 体积/体积, 6ml) 中对该双乙酸酯 (300 mg, 1.6mmol) 进行对映体选择性二羟基化。24 小时后, 用固体 Na_2SO_3 终止反应, 并继续搅拌 30 分钟。用乙酸乙酯重复萃取 (6×12 ml) 和急骤层析 (乙酸乙酯), 获得一种仅含 2-C-甲基-D-赤藓醇双乙酸酯的混合物 (312 mg,

88%产率)(得自部分分子内转酯化作用)。在有碱性大孔树脂 Amberlyst A-26(OH-型)(每毫摩尔 150 毫克)的存在下在甲醇(30 ml)中于室温过夜,进行定量脱乙酰化(Reed 等人,1981)。在对树脂进行过滤和对溶剂进行蒸发后,就可直接获得纯的 2-C-甲基-D-赤藓醇(190 mg, 75%的总产率)。

实施例 2: 大肠杆菌 dxr 基因的定点标记插入诱变

使用从野生型大肠杆菌菌株 W3110 中分离的基因组 DNA (Kohara 等人, 1987) 以及引物 P1 (5'-CTCTGGATGTCATATGAAGCAACTC-3' (SEQ ID NO: 3), 其中下划线的 ATG 相当于 dxr 基因的翻译起始密码子) 和 P2 (5'-CCGCATAACACCGCCAACC-3' (SEQ ID NO: 4), 其位于 yaeS 基因的 3' 侧翼区域), PCR 扩增从 dxr 基因 5' 区至 yaeS 基因 3' 侧翼区域的延伸区。PCR 反应混合物的终体积为 50 μ l, 其中含有 DNA 模板(100ng)、每种引物 0.5 μ M、每种脱氧核苷三磷酸 200 μ M、pH 调至 8.8 的 Tris-HCl 20mM、MgSO₄ 2mM、KCl 10mM、(NH₄)₂SO₄ 10mM、BSA 0.1mg/ml 和 Triton X-100 0.1%。将样品用矿物油覆盖, 在 94°C 温育 3 分钟, 然后冷却至 80°C。加入 Pfu DNA 聚合酶(1.25 个单位, Stratagene), 然后将反应混合物温育 30 个循环, 其中每个循环为: 94°C 45 秒、59°C 45 秒和 72°C 10 分钟, 循环完成后, 接最后一步 72°C 10 分钟。扩增后, 根据制造商的说明书方法对 PCR 产物进行 3' 末端加腺嘌呤, 然后将腺苷酰化的产物克隆到 pGEM-f 载体(Promega)中, 以生成质粒 pMJ1。将存在于质粒 pCAT19(Fuqua, 1992)中的 CAT(氯霉素乙酰转移酶)基因用 PstI 和 XbaI 消化切割出来, 用 T4 DNA 聚合酶处理, 然后通过钝端连接克隆到 dxr 基因中的独有 AsuII 位点上(在用 T4 DNA 聚合酶处理后), 得到质粒 pMJ2。使用了限制酶图谱以鉴定这种克隆, 其中的 CAT 基因与 dxr 基因处于同一方向。将从质粒 pMJ2 中切割下来的 SpeI-SphI 片段亚克隆至质粒 pBR322 的 NheI-SphI 位点上, 构建出质粒 pMJ3。用 PstI 消化使质粒 pMJ3 线性化, 用小牛肠碱性磷酸酶(GibcoBRL)温育, 用琼脂糖凝胶电泳纯化。用 2 μ g 纯化的线性质粒 pMJ3 转化大肠杆菌菌株 JC7623(Winans 等人,

1985)。将转化细胞涂于补充有 2mM 2-C-甲基-D-赤藓醇 (ME) 以及氯霉素 (17 μ g/mL) 的 LB 平板 (Ausubel 等人, 1987) 上。将既显示氯霉素抗性又显示 ME 营养缺陷的克隆选择出来供进一步研究。使用引物 P3 (5' -GCACACTTCCACTGTGTGTG-3' (SEQ ID NO: 5), 其位于 *frr* 基因的 5' 区域) 和 P2, 通过 PCR 检测 CAT 基因已插入到 *dxr* 基因中。将这些克隆中的一个 (被指定为菌株 JC7623 *dxr*: CAT), 用于互补实验。

实施例 3: cDNA 末端快速扩增 (RACE)

为了鉴定推定为编码 1-脱氧-D-木酮糖 5-磷酸还原异构酶 (DXR) 的同源物的植物核酸序列, 用最近克隆的大肠杆菌 DXR (Takahashi 等人, 1998) 的完整氨基酸序列作为查询序列, 利用 TBLASTN 程序搜索国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 的非冗余 (Non-Redundant) 数据库。在该查询序列与由拟南芥 (*A. thaliana*) 基因组克隆 MQB2 (收录号 ABOO9053) 的 7 个推测的外显子所编码的氨基酸序列之间发现了相当水平的相同性 (40-64%)。

为了证实与推定的拟南芥 DXR 基因相对应的 mRNA 序列的存在, 用克隆 MOB2 的核苷酸序列 (从核苷酸 29247 延伸至 31317) 作为查询序列, 利用 BLASTN 程序搜索 NCBI 的 EST 数据库 (dbEST)。发现了两个拟南芥 EST 克隆 (12OE8T7 和 65F11XP3', 收录号分别为 T43949 和 AA586087), 它们含有与查询序列的不同区域相同的核苷酸序列。对 cDNA 插入物的测序表明, 这两个克隆有重叠。最长的 cDNA 中含有一个开放读框, 它编码一个 329 个残基的多肽, 其与大肠杆菌 DXR 的 C-末端区域有 41.6% 的相同性 (相似性 53.2%), 这就表明这两个 cDNA 编码了被认为是拟南芥酶的截短形式。

按 Dean 等人 (1985) 的描述, 纯化光照生长 12 天的拟南芥 (Columbia 变种) 幼苗中的总 RNA。采用 Life Technologies/Gibco BRL 的 5' -RACE-系统 (2.0 版本), 根据提供商的指示, 进行 cDNA 末端快速扩增 (RACE)。使用 1 μ g RNA 样品作为模板以及寡核苷酸 DXR-GSP1 (5' -ATTCGAACCAGCAGCTAGAG-3' (SEQ ID NO: 6),

与 SEQ ID NO: 1 中显示的核苷酸+767 至+786 互补) 作为特定的下游引物, 合成 cDNA 的第一链。在对 cDNA 进行纯化和同聚物加尾后, 进行两个嵌套式 PCR 反应。在第一个 PCR 反应中, 特定的下游引物为寡核苷酸 DXR-GSP2 (5' -CCAGTAGATCCAACGATAGAG-3' (SEQ ID NO: 7), 与 SEQ ID NO: 1 中显示的核苷酸+530 至+550 互补), 上游引物为寡核苷酸 5' -RACE-AAP (试剂盒中提供)。在第二个 PCR 反应中, 特定的下游引物为寡核苷酸 DXR-GSP3 (5' -GGCCATGCTGGAGGAGGTTG-3' (SEQ ID NO: 8), 与 SEQ ID NO: 1 中显示的核苷酸+456 至+475 互补), 上游引物为寡核苷酸 AUAP (试剂盒中提供)。在这两个 PCR 反应中, 扩增过程均起始于样品的变性 (在 94°C 下 3 分钟), 冷却至 80°C, 然后加入 Taq DNA 聚合酶。第一个 PCR 反应的混合物温育 15 个循环, 每个循环为 94°C 30 秒、55°C 30 秒和 72°C 1 分钟, 循环完成后, 接最后一步 72°C 5 分钟。将所得到的样品按 1 比 10 稀释于第二个 PCR 反应混合物中, 然后温育 30 个循环, 每个循环为 94°C 30 秒、61°C 30 秒和 72°C 1 分钟, 最后一步为 72°C 5 分钟。用琼脂糖凝胶电泳对最终的扩增产物进行纯化, 然后克隆到质粒 pBluescript SK+ 上并进行测序 (SEQ ID NO: 1)。

实施例 4: 从拟南芥中克隆 1-脱氧-D-木酮糖 5-磷酸还原异构酶的 cDNA

为了确定推定的 DXR 基因的 5' 区域, 使用 RACE 技术对相应的转录起始位点进行定位。在对大肠杆菌的 DXR 与从拟南芥基因组克隆推导出的氨基酸序列进行比对的基础上, 设计了引物。在 SEQ ID NO: 2 中提供了从拟南芥菜属的 dxr 核酸序列 (SEQ ID NO: 1) 推导出的氨基酸序列。用拟南芥籽苗的 RNA 作为模板, 用寡核苷酸 DXR-GSP1 作为引物, 合成了 cDNA 的第一条链。其中的寡核苷酸与 SEQ ID NO: 1 中显示的基因组序列+767 位至+786 位间的区域互补。随后, 进行两个嵌套式 PCR 反应, 以扩增 mRNA 的 5' 末端。用于第一及第二个嵌套式 PCR 反应的下游特定引物分别与从+530 位至+550 位 (引物 DXR-GSP2) 和+456 位至+475 位 (引物 DXR-GSP3) 的区域互补。对与主要扩增产物对应的 4 个克隆进行了测序并发现它们都具有相同

的 5' 末端, 该末端与 SEQ ID NO: 1 中显示的基因组序列的+1 位腺嘌呤相对应。

通过两个连续 PCR 反应, 从源自拟南芥 (Columbia 变种) 的细胞系 T87 悬浮液的 cDNA 文库中, 扩增含有拟南芥菜属完整的 DXR 编码序列的 cDNA。将一份文库用乙醇进行沉淀, 然后再悬浮于水中。第一个 PCR 反应的混合物终体积为 25 μ l, 其中含有: DNA 模板 (相当于 cDNA 文库的 4×10^5 个空斑形成单位), 上游引物 DXR-34 (5' - CAAGAGTAGTAGTGCGGTTCTCTGG-3' (SEQ ID NO: 9), 与 SEQ ID NO: 1 中显示的序列+34 至+58 相对应) 0.5 μ M, 下游引物 DXR-E2 (5' - CAGTTTGGCTTGTTTCGGATCACAG-3' (SEQ ID NO: 10), 与 SEQ ID NO: 1 中显示的核苷酸序列+3146 至+3169 互补) 0.5 μ M, 每种脱氧核苷三磷酸 200 μ M、pH 调至 8.8 的 Tris-HCl 20mM、MgSO₄ 2mM、KCl 10mM、(NH₄)₂SO₄ 10mM、BSA 0.1mg/ml 和 Triton X-100 0.1%。将样品用矿物油覆盖, 在 94°C 下温育 3 分钟, 然后冷却至 80°C。加入 Pfu DNA 聚合酶 (1.25 个单位, Stratagene), 然后将反应混合物温育 35 个循环, 每个循环为 94°C 30 秒、55°C 40 秒和 72°C 6.5 分钟, 循环完成后, 接最后一步 72°C 15 分钟。用水将反应混合物按 1: 10 稀释, 取出 5 μ l 用作第二个 PCR 的模板, 进行第二个 PCR 的条件, 除了将反应体积增加到 50 μ l 和将循环数减至 15 个以外, 均与所描述的前面的扩增条件相同。用琼脂糖凝胶电泳对扩增产物进行纯化, 然后克隆到质粒 pBluescript SK+上。将所得质粒命名为 pDXR-At。

这样, 使用 DXR-34 和 DXR-E2 (分别与基因组序列的从+34 位至+58 位以及+3146 位至+3169 位的延伸区域相对应) 为引物, 从 cDNA 文库中通过 PCR 获得了编码拟南芥的完整 DXR 的 cDNA 克隆。通过 DNA 测序, 证实了所扩增的 cDNA 的相同性。对 cDNA 与基因组序列所进行的比对表明, 拟南芥 DXR 基因含有 12 个外显子和 11 个内含子, 其延伸区域为 3.2Kb (SEQ ID NO: 1)。

所克隆的 cDNA 编码一个 477 个氨基酸残基的蛋白质, 其推定的分子量为 52 kDa。对拟南芥与大肠杆菌的 DXR 所进行的比对 (图 1) 显示, 这个植物酶有一个具有质体转移肽的典型特征的 79 个残基的 N-末端延伸 (von Heijne 等人, 1989)。这两个蛋白质显示 42.7% 的相

同性（54.3%的相似性）。

实施例 5：表达型构建体的制备

为了在大肠杆菌中表达拟南芥 DXR，从质粒 pDXR-At 中 PCR 扩增 DXR cDNA 的编码氨基酸残基 81-477 的区域，并将其克隆到经过修饰的质粒 pBAD-GFPuv (Clontech) 中。在这个质粒中，表达是由 P_{BAD} 启动子驱动的，该启动子可被阿拉伯糖诱导和被葡萄糖抑制。首先，对质粒 pBAD-GFPuv 进行修饰，依照 Kunkel 等人的方法 (Kunkel 等人, 1987) 采用定点诱变将位于 pBR322ori 和 araC 编码区之间的 NdeI 位点 (4929-4931 位) 除去。将寡核苷酸 pBAD-mut1 (5' -CTGAGAGTGCACCATCTGCGGTGTGAAATACC-3' (SEQ ID NO: 11)) 用作诱变引物。将所生成的质粒指定为 pBAD-Mi。接着，通过 PCR 将 NdeI 和 EcoRI 限制位点引入到拟南芥 DXR cDNA 的合适部位中，其中使用质粒 pDXR-At 作为模板，使用寡核苷酸 5' -MVKPI (5' -GGCATATGGTGAAACCCATCTCTATCGTTGGATC-3' (SEQ ID NO: 12)，与在 SEQ ID NO: 1 中列出的核苷酸序列 +522 至 +544 互补，其中下划线序列含有 NdeI 位点) 和 DXR-END (5' -ACGAATTCATTATGCATGAACTGGCCTAGCACC-3' (SEQ ID NO: 13)，与在 SEQ ID NO: 1 中列出的核苷酸序列 +2997 至 +3018 互补，其中下划线序列含有 EcoRI 位点) 作为诱变引物。用 NdeI 和 EcoRI 消化 PCR 扩增产物并将其克隆到用相同的限制酶消化的质粒 pBAD-M1 中。这就使得质粒 pBAD-M1 中的 GFPuv 编码序列被拟南芥 DXR 的相应的编码序列替代。将所得质粒 (被指定为 pBAD-DXR) 引入到菌株 XL1-Blue 中。在互补实验中，编码 GFPuv 的质粒 pBAD-M1 被作为对照使用。

实施例 6：拟南芥 dxr 的分析

通过互补实验分析一株其所携带的 dxr 基因中有一处被破坏的大肠杆菌 (菌株 JC7623 dxr.: CAT) (见实施例 2)，已建立了所克隆的拟南芥 DXR 的功能。菌株 JC7623 dxr.: CAT 的生长需要 2-C-甲基-D-赤藓醇 (ME)。在互补实验中，我们使用了拟南芥 DXR 的从氨基酸 81

至 477 所延伸的一段区域（见 SEQ ID NO: 2），它不包括推定的质体转移肽。将合适的 cDNA 片段克隆到质粒 pBAD-GFPuv 的一个衍生物中，并处于 PBAD 启动子的控制之下，然后将所得到的质粒（pBAD-DXR）引入到 JC7623 dxr.-CAT 菌株中。从 PBAD 启动子的表达可被阿拉伯糖诱导和被葡萄糖抑制。用阿拉伯糖诱导可以使含有质粒 pBAD-DXR 的菌株 JC7623 dxr.-CAT 在没有 ME 存在的情况下生长，而在有葡萄糖存在的情况下没有观察到该菌株的任何生长。携带控制质粒 pBAD-M1 的菌株 JC7623 dxr.,: CAT 在有阿拉伯糖存在、缺乏 ME 的培养基上并不生长。携带质粒 pBAD-DXR 或者 pBAD-GFPuv 的菌株 JC7623 dxr.-CAT 在含有 ME 的培养基上生长。这些结果明确无误地证实了所克隆的拟南芥 cDNA 编码一个功能性 DXR。

在本发明中所提到的所有出版物和专利申请，都表示本发明所属于的本领域技术人员的技术水平。本文中被引用作为参考的所有出版物和专利申请，都是以特定的和独立的方式引用作为参考。

尽管前面已通过例证和实施例对本发明进行了一些详细描述，以便对本发明有一个清楚的理解，但是很显然，在所附的权利要求书的范围内，还可以对本发明进行某些变化和修饰。

序列表

<110> 卡尔根尼有限公司 (Calgene LLC)

<120> 编码参与类异戊二烯合成的蛋白质的核酸序列

<130> 17142/00/WO

<150> US 60/129,899

<151> 1999-04-15

<150> US 60/146,461

<151> 1999-07-30

<150> PCT/US00/10367

<151> 2000-04-14

<160> 13

<210> 1

<211> 3400

<212> DNA

<213> 拟南芥菜属种

<400> 1

```

cttgttacta aatgctcagc gaaatcttta aaaaatgaca aaaatctgtt ggggtaccatt      60
caaatccaga ttccctttctt atcatcatct ctctctctca cactgtttat ctgattcgtc      120
ttctctgata atcaagagta gtagtgcggt tctctggaaa atattcgatt tttaaaagac      180
tctgatgatg acattaaact cactatctcc agctgaatcc aaagctatct ctttcttgga      240
tacctccagg tccaatccaa tccctaaact ctccaggtttc ttcttcttcc tctcttcttt      300
cctcctcctt ggtcaactct cttttcgatt aaagtigcaa actttcatta gttgtcttag      360
gctcttgtga atttctctat ctaggtaatc tggtatttct tcaattcgat tttttttggt      420
ttgcttttag tcgtagagtt ttaaatttta catctttgga gtgtttcaca ggtgggttta      480
gtttgaggag gaggaatcaa gggagagggt ttggaaaagg tgtaagtgt tcagtgaag      540
tgcagcagca acaacaacct cctccagcat ggcctgggag agctgtccct gaggcgcctc      600
gtcaatcttg ggatggacca aaacccatct ctatcgttgg atctactggt tctattggca      660
ctcaggtttt atttcgatta aggcatatt gtgcagttct tgagtatgac cagactttaa      720
gtttgtctta tgaatgacta gactcataga agaattgatat ttttttctta ctgagttatt      780
gttgcatcat ttttatcgac aagaacttcc attttgaga cattggatat tgtggctgag      840
aatcctgaca aattcagagt tgtggctcta gctgctggtt cgaatgttac tctacttgct      900
gatcaggtaa gttggcttca ttgttaaaaa aattagtatt gagtctctcc aatttgtcat      960
tcagaccact tggaaattcag ttttaattctc agttcagttg tagtatcata agcaagatag     1020
tattaactcg ttatgtatca gatcaaacca gagaaatcag gttctggttt aggcttttgc     1080
ttctgcaatc tcaagaaatc tctatagtat ggttctgtga ttctattttg aatggtggca     1140
ggtaaggaga ttttaagcctg catttggttg ttttagaaac gagtcaactga ttaatgagct     1200
taaagaggct ttagctgatt tggactataa actcgagatt attccaggag agcaaggagt     1260
gattgagggt agttcatttg ttagttttga ttgtagtgta gataggtttt tacttattat     1320
gttcatcaac aggttgcccg acatcctgaa gctgtaaccg ttgttaccgg aatagtaggt     1380

```

```

tgtgcgggac taaaggtata tactctaatt ttttgttatt aaaccttatt aagaggatat 1440
gaaaaaagaa agttgcagat gataaagctt gttgcttatt ttactgcag cctacggttg 1500
ctgcaattga agcaggaaag gacattgctc ttgcaaacia agagacatta atcgcaggtg 1560
gtcctttcgt gcttccgctt gccaaacaaac ataatgtaaa gattcttccg gcagattcag 1620
aacattctgc catatttcag gtatcacaaa tcacatagaa ttaagtacct caactttcat 1680
attgagttca gcgttgggtc taatgcaagt tcaacctctg gcaatttgag tgaaaaatct 1740
tcttttatgt tctctagtgt attcaagggt tgccgtgaagg cgctctgcgc aagataatct 1800
tgactgcac tgggtggagct tttaggtttg tttcgatatt cttctctctc tgcatagact 1860
ttttttcttc tcaattctcg tttgggttaat ggaaactttt cactggattt tgaaaaagg 1920
attggcctgt cgaagagcta aaggaagtta aagtagcgga tgcgttgaag catccaaact 1980
ggaacatggg aaagaaaatc actgtggact ctgctacgct tttcaacaag gtttaagatta 2040
ttttctccta aggttaaact ctgattttga aaataccttt gatcaaggta gatgagttct 2100
tgattttttg aaacagggtc ttgaggtcat tgaagcgcat tttttgttg gagctgagta 2160
tgacgatata gagattgtca ttcatccgca aagtatcata cattccatga ttgaaacaca 2220
ggtcttgctg aaacattact aactaaatta ttatttttcc ggttttaaaa aaataactgt 2280
ataacatgta tttgttttgt tccacaggat tcatctgtgc ttgctcaatt gggttggcct 2340
gatatgcgtt taccgattct ctacaccatg tcatggcccg atagagttcc ttgttctgaa 2400
gtaacttggc caagacttga cttttgcaag taagctaacc acatttatat actctctgtt 2460
tatcaagtgt gaagctaagc ttagttgaaa attttaatta tcaccaagaa aagtcccca 2520
atcttgtttt cagtttgggt ttaggttgtt tagataagat aaaaaatgaa accgaatcgg 2580
tcttcgggtt gggtttgcaa ttgggttatt tgctactgtt ttgggtgtgga tcagttaaac 2640
tgggttagga ccactgcctt atctatcagc attcagcacc taaaaccaa agttgtttac 2700
aattgtggat tttggcagac tcggttcatt gactttcaag aaaccagaca atgtgaaata 2760
cccatccatg gatcttgctt atgctgctgg acgagctgga ggcacaatga ctggagttct 2820
cagcgccgcc aatgagaaaag ctgttgaaat gttcattgat gaaaagtaag aattatttt 2880
cagttttgag catctcaatg aagttcttga tacgaatcac aattgtttat attctcactt 2940
ttgtttacag gataagctat ttggatatct tcaaggttgt ggaattaaca tgcgataaac 3000
atcgaaacga gttggtaaca tcaccgtctc ttgaagagat tgttcactat gacttgtggg 3060
cacgtgaata tgccgcgaat gtgcagcttt cttctgggtc taggccagtt catgcatgaa 3120
gaattgggtt ttggaagaac ataaggaagc ttctgaggaa atgttgaaag aagattagtg 3180
tagagaatgg ggtactactt aatagcgttt ttggcaagga ttatggattg ttagctaat 3240
ttatctgtga tccgaacaag ccaaactgat aatttgaac catttttacc aataaaaccg 3300
agcttaattg tttcacatta tatgattaat tacattcatc taagggttct tgaaaagcct 3360
ctgagcttca tgagtagagt tcgcatctcc tgttgcgtc 3400

```

<210> 2

<211> 476

<212> PRT

<213> 拟南芥菜属种

<400> 2

```

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 Asn 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	Leu	Glu	Ile	Ile	Pro	Gly
145					150					155					160
Glu	Gln	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
Asn Val Gln Leu Ser Ser Gly Ala Arg Pro Val His Ala		
465	470	475

<210> 3
 <211> 25
 <212> DNA
 <213> 人工序列
 <220>
 <223> 合成寡核苷酸

<400> 3
 ctctggatgt catatgaagc aactc 25

<210> 4
 <211> 19
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 4
 ccgcataaca ccgccaacc 19

<210> 5
 <211> 20
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 5
 gcacacttcc actgtgtgtg 20

<210> 6
 <211> 20
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 6
 attcgaacca gcagctagag 20

<210> 7
 <211> 21

<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 7
ccagtagatc caacgataga g 21

<210> 8
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 8
ggccatgctg gaggaggttg 20

<210> 9
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 9
caagagtagt agtgcggttc tctgg 25

<210> 10
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 10
cagtttggct tggtcggatc acag 24

<210> 11
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 11
ctgagagtgcc accatctgcg gtgtgaaata cc 32

<210> 12
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 12
ggcatatggt gaaacccatc tctatcggtg gatc 34

<210> 13
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 13
acgaattcat tatgcatgaa ctggcctagc acc 33

```

At      MMTLNSLSPA ESKAISFLDT SRFNPIPKLS GGFSLRRRNQ GRGFGKGVKC

At      SVKVQQQQQP PPAWPGRAVP EAPRQSWDGP KPISIVGSTG SIGTQTLDIV
E col           M KQLTILGSTG SIGCSTLDVV

At      AENPDKFRVV ALAAGSNVTL LADQVRRFKP ALVAVRNESL INELKEALAD
e col    RHNPEHFRVV ALVAGKNVTR MVEQCLEFSP RYAVMDDEAS AKLLKTMLQQ

At      LDYKLEIIPG EQGVIEVARH PEAVTVVTGI VGCAGLKPTV AAIEAGKDIA
e col    QGS RTEVLSG QQAACDMAAL EDVDQVMAAI VGAAGLLPTL AAIRAGKTIL

At      LANKETLIAG GPFVLPLANK HNVKILPADS EHS AIFQCIQ -----
e col    LANKE SLVTC GRLFM DAVKQ SKAQLLPVDS EHNAIFQSLP QPIQHN LGYA

At      GLPEGALRKI ILTASGGAFR DWPVEKLKEV KVADALKHPN WNMGKKITVD
e col    DLEQNGV VSI LLTGSGGPFR ETPLRDLATM TPDQACRHPN WSMGRKISVD

At      SATLFNKGLE VIEAHYLFGA EYDDIEIVIH PQSIIHSMIE TQDSSVLAQL
e col    SATMMNKGLE YIEARWLFNA SASQMEVLIH PQSVIHSMVR YQDGSVLAQL

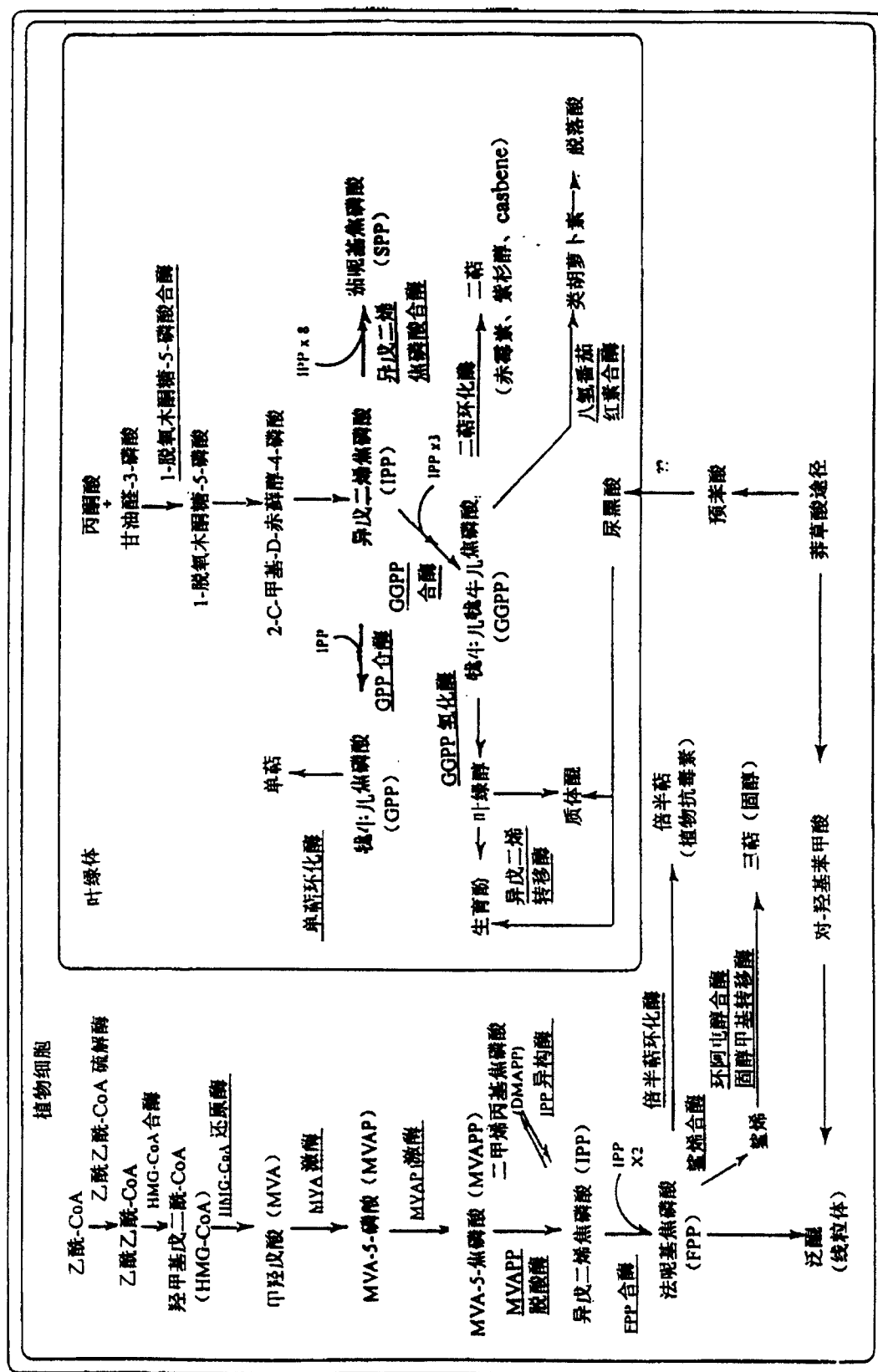
At      GWPDMRLPIL YTMSWPDRVP CSEVTWPRLD LCKLGSLTFK KPDNVKYPSM
e col    GEPDMRTPIA HTMAWPNRVN SGV---KPLD FCKLSALIFA APDYDRYPCL

At      DLAYAAGRAG GTMTGVLSAA NEKAVEMFID EKISYLDIFK VVELTCDKHR
e col    KLAMEAFEQG QAATTALNAA NEITVAAFLA QQIRETDIAA LNLSVLEK--

At      NELVTSPSLE EIVHYDLWAR EYAANVQLSS GARPVHA
e col    MDMREPQCVD DVLSSVDANA REVARKEVMR LAS

```

图 1



2
图