[19]中华人民共和国国家知识产权局

[51]Int. Cl[7]
C08H 5/02
C08L 97/00 D21C 5/00
C09B 69/10 C08F 2/00

[12]发明专利申请公开说明书

[21]申请号 01117111.1

[43]公开日 2001年12月5日

[11]公开号 CN 1324882A

[22]申请日 1994.9.13 [21]申请号 01117111.1
分案原申请号 94193815.8
[30]优先权
[32]1993.9.17 [33]US [31]08/122,230
[32]1993.9.17 [33]US [31]08/122,827
[32]1993.12.3 [33]US [31]08/162,827
[32]1993.12.22 [33]US [31]08/172,331
[71]申请人 诺沃挪第克公司
地址 丹麦巴格斯瓦尔德
共同申请人 诺沃挪第克生物技术公司
[72]发明人 J·A·瓦雷斯尼尔
B·E·克里斯蒂森
P·施奈德

[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 李 瑛

权利要求书1页 说明书48页 附图页数21页

[54]发明名称 纯化的pH中性丝核菌属漆酶及编码该酶的核酸

[57]摘要

本发明涉及分离的、含有编码在中性或碱性pH下具有最适活 性的茄丝核菌漆酶之序列的核酸片段及由其编码的漆酶蛋白质。

ISSN1008-4274

权 利 要 求 书

1. 一种在溶液中聚合木质素或硫酸木质素底物的方法，包含将该底物与在 pH 为大约 6.0－8.5 之间发挥最佳功能的丝核菌属漆酶接触。

2. 一种在牛皮纸浆中进行原位解聚的方法，包含将该纸浆与在 pH 为大约 6.0－8.5 之间发挥最佳功能的丝核菌属漆酶接触。

3. 一种氧化染料的方法，包含将该染料与在 pH 为大约 6.0－8.5 之间发挥最佳功能的丝核菌属漆酶接触。

4. 一种聚合酚类化合物的方法，包含将该酚类化合物与在 pH 为大约 6.0－8.5 之间发挥最佳功能的丝核菌属漆酶接触。

说　　明　　书

纯化的 *pH* 中性丝核菌属漆酶及编码该酶的核酸

相关申请

本申请是共同未决的美国系列号 08/122，230，08/122，827 和 08/162827 的部分延续，引用其全文的内容以供参考。

发明领域

本发明涉及分离的编码真菌氧化还原酶的核酸片段及由此产生的纯化的酶。更具体地说，本发明涉及编码在中性 *pH* 下发挥功能的酚氧化酶（特别是漆酶）的核酸片段。

发明背景

漆酶（苯二酚：氧氧化还原酶）是催化酚醛塑料氧化的含多个铜离子的酶。漆酶介导的氧化导致以合适的酚底物产生芳氧基一基团中间体；所产生的中间体最终偶联而提供一种二聚体一寡聚体和多聚体反应产物的组合物。这类反应在自然界导致黑色素，生成碱，毒素，木质素和腐殖酸形成的生物合成途径中是重要的。漆酶可由很多种真菌产生，包括子囊菌类（如曲霉属、链孢霉属和柄孢壳菌属）、半知菌类葡萄孢属和担子菌类（如金线菌属，屈孔菌属，香菇属，灰侧耳菌属，栓菌属）以及丝核菌属的有性生殖菌形式。漆酶表现出大范围的底物特异性，每种不同的真菌漆酶通常只能按其氧化酚底物的能力以定量的方式来互相区别。由于具有底物多样性，人们普遍发现了漆酶的许多有潜力的工业应用。其中有木质素修饰、纸强化、在去污剂中的染料转移抑制、酚聚合、汁液加工，酚树

脂生产和废水处理。

尽管催化能力相似，但不同真菌种类产生的漆酶具有不同的温度和 *pH* 最适值，这些也可以根据特异性底物来区别。已分离出许多这类真菌漆酶，其中一些基因已被克隆。例如，*Choi* 等（*Mol. Plant—Microbe Interactions* 5:119—128,1992）描述了编码栗树枯萎病真菌（*Cryphonectria parasitica*）漆酶的基因的分子特征和克隆。*Kojima* 等（*J. Biol, Chem.* 265:15224—15230,1990; *JP*2—238885）提供了对白色腐烂病担子菌毛革盖菌漆酶两种等位形式的描述。*Germann* 和 *Lerch*（*Experientia* 41 :801,1985; *PNAS USA* 83:8854—8858, 1986）报道了粗糙链孢霉漆酶基因的克隆和部分测序。*Saloheimo* 等（*J. Gen. Microbiol.* 137 :1537—1544 ,1985 ; *WO*92/01046）公布了对真菌射脉菌漆酶基因的结构分析。然而，实际上所有已知的真菌漆酶都在酸性 *pH*（例如：在 *pH*3.0 和 6.0 之间）下发挥最佳功能，且在中性和碱性 *pH* 下一般无活性。由于许多前述有潜力的工业方法优选在中性或碱性 *pH* 下进行，在这类方法中大多数真菌漆酶利用效果较差。因此，在许多重要商用方法的应用中，可得到的真菌漆酶是不充足的。

这一常规的一个例外是丝核菌属的某些种类所产生的胞外漆酶。*Bouag* 等报道了 *Rhizoctonia praticola* 产生的最适 *pH* 大约为 7.0 的漆酶。这种类型的漆酶与以前已知的漆酶相比在需要中性 *pH* 的工业方法上更为有用。然而，这种 *R. Praticola* 酶既未纯化，也未进一步鉴定，而且到现在为止没有任何其它具有该特征的漆酶被纯化或鉴定。而且，尽管已分离出了其它的漆酶基团（如上所述），但这些基因编码的酶在酸性 *pH* 下发挥最佳功能。因此还不能

达到对 *pH* 中性或碱性漆酶的重组生产和足够的商用产量，因为该酶本身和编码该酶 的漆酶基因以前尚未被分离和/或纯化及测序过 。本发明现在提供了对这些问题的解决方法。

本发明概述

本发明涉及分离的含有编码在 *pH*6.0 到 8.5 之间发挥最佳功能的丝核菌属漆酶的核酸序列的核酸片段。“发挥最佳功能”是指酶在 *pH* 范围为大约 6.0 到 8.5 之间时表现出明显的（即 ，至少是最大值的大约 30% ，优选至少大约 50%，最优选从 50%到最大值）活性，以诸如丁香醛连氮 、*ABTS*、2，6—二甲氧基酚或 4 *antiaminopyrine*＋*N*—乙基—*N*—磺基丁 基—*m*—甲苯胺的底物的一种或多种标准漆酶试验中的活性来确定。本发明的漆酶优选的底物是丁香醛连氮。在一个优选的实施例中 ，漆酶是茄丝核菌漆酶。本发明还涉及由新核酸序列编码的基本上纯的漆酶。“基本上纯的”指基本上（即：≥90%）无其它非漆酶蛋白质的漆酶。

为了便于生产新的漆酶，本发明还提供了包含所要求保护的核酸片段的载体和宿主细胞 ，该载体和宿主细胞在漆酶的重组生产中有用处 。该核酸片段与触指导漆酶蛋白质在选择的宿主细胞中表达的转录和翻译信号以可操纵 的方式相连。优选的宿主细胞是真菌细胞，最优选的是曲霉属。本发明的漆酶的重组生产通过在适于漆酶蛋白质表达的条件下培养用于本发明的核酸片段转化或转染的宿主细胞或其子代并从培养物中回收漆酶蛋白质来实现。

本发明的漆酶在许多需要氧化酚醛塑料的工业方法中有用。这些方法包括木质素修饰 ，汁液加 2，酚聚合和酚树脂的生产。在优选的实施例中 ，本发明的酶用于为达到最大效率而需要中性或略带

碱性 *pH* 的方法中。

图 1 显示了 *RSlac*1 的核苷酸和氨基酸顺序 。核苷酸序列中的小写字母表示内含子位置。

图 2 显示了 *RSlac*2 的核苷酸和氨基酸顺序 。核苷酸序列中的小写字母表示内含子位置。

图 3 显示了质粒 *pMWR*—1 的限制性图谱。

图 4 显示了 *RSlac*3 翻译区的核苷酸和氨基酸序列。

图 5 显示了 *RSlac*1 的丁香醛连氮氧化酶活性(90*mM* 缓冲液，20*μm* 丁香醛连氮，20℃)。

图 6 显示了 *RSlac*2 的丁香醛连氮氧化酶活性(93*mM* 缓冲液，20*μm* 丁香醛连氮，20℃)。

已报道了生产漆酶的丝核菌属的某些种类；因此，起始研究集中于鉴定在各种茄丝核菌属分离物中这些酶的存在。培养样品并以 *ABTS* 方法(在下面描述)定期分析上清液中漆酶的存在。在全部丝核菌属培养物中观察到漆酶 。收获的漆酶经电泳分离并用 *ABTS* 染色，一种分离物 *RS*22 产生具有碱性 *pI* 的漆酶，选来进行进一步的研究。

后续研究集中于对来自 *RS*22 的酶进行纯化和鉴定 。简单地说，过滤发酵培养基并用具有以大约 10 ,000 倍浓缩的膜的 *UF* 进行浓缩 。在 *pH*4.5 下在乙酸缓冲液中进行第一步离子交换层析步骤，使用 *NaCl* 进行洗脱步骤 。然后洗脱物经过滤并再进行层析，用 *NaCl* 梯度洗脱。汇集活性部分以备进一步研究。

首先对由此分离和纯化的完整蛋白质(下文称为 *RSlac*3)进行部分测序，获得的 *N* 末端顺序如下：

AVRNYKFDIKNVNVAPDGFQRPIVSV(*SEQ. ID. NO.* :5)

用赖氨酸或谷氨酸特异性蛋白酶对该蛋白质进行进一步消化，从该蛋白质获得的其它的肽具有下列序列，可与毛革盖菌中的序列进行序列对比：

肽 1：

SQYVDGLRGPLVIYDPDDDH(*SEQ. ID. NO*:6)

肽 2：

GLALVFAEAPSQIRQGVQSVQPDDA(*SEQ. ID. NO*:7)

肽 3：

SRYBVBBASTVVMLEBWYHTPAXVLE(*SEQ. ID. NO*:8)

肽 4：

SLGPTPNYVNPXIRDVVRVGGTTVV(*SEQ. ID. NO*:9)

还发现了下列多肽，但与革盖菌属序列不相关。

肽 5：

IRYVGGPAVX(*N*?)*RSVI*(*SEQ. ID. NO*:10)

肽 6：

ILANPA(*SEQ. ID. NO*:11)

肽 7：

YEAPSLPT(*SEQ. ID. NO*:12)

在上面序列中，*B* 代表天冬氨酸或天冬酰胺的，*X* 代表未确定的残基。

为了起动丝核菌属漆酶基因的筛选，制备了茄丝核菌基因组文库。用限制性酶 *Sau*3*A* 部分消化总 *DNA*，在琼脂糖凝胶上光电泳以分离大小，在 8 和 21*kb* 之间的 *DNA* 片段。分离的片段以

BamHI 末端连接到入噬菌体 *EMBL*3 臂上，所得的噬菌体进行体外包装。这些噬菌体用作文库以在大肠杆菌中产生 170,000 噬菌斑的文库并扩增 100 倍备用。

为了形成用于分离茄丝核菌漆酶基因的探针，分析了 5 个已知漆酶的蛋白质序列以确定相同序列，根据得到的相同序列构建了 2 个简并寡聚核苷酸（*Choi et al.* 出处同上；*Germann and Lerch*，出处同上；*saloheimo etal* 出处同上；*Kojima et al*，出处同上）。将这些寡聚核苷酸与茄丝核菌基因组 *DNA* 混合，使用 *Taq* 聚合酶链式反应（*PCR*）扩增出 220 核苷酸片段的 *DNA* 片段。然后将 220 核苷酸片段克隆进质粒载体中。

PCR 片段用作探针以便以扩增的基因组文库中筛选 25，000 个噬菌斑。将这些筛选的阳性克隆分成随后经 *DNA* 序列分析显示为编码 2 种不同漆酶基因（*RSlac*1 和 *RSlac*2）的两种类型。图 1 和图 2 分别列出了每个基因的核苷酸序列（*SEQ ID. NO*：1 和 3）和预测的每种蛋白质的氨基酸序列（*SEQ ID*，*NO*：2 和 4）。两序列之间的同源性大约为 63%。与来自毛革盖菌，辐射射脉菌，构巢曲霉，*Cryphonectria parasitica* 和粗糙链孢霉的已知漆酶序列相比，*RS* 漆酶表现出大约在 30—40%之间的同源性。将 2 种编码序列分别与米曲霉高峰氏淀粉酶转录和翻译信号以可操纵的方式相连克隆进表达载体中（见图 3）。分别将 2 种漆酶表达载体转化进米曲霉和黑色曲霉宿主细胞中，筛选存在漆酶的宿主细胞。

为了分离 *RSlac*3 基因，以在菌香胺存在的条件下生长的茄丝核菌菌丝体中纯化 *ployA RNA*。该 *RNA* 用作 *cDNA* 合成的模板。分离 *cDNA* 并收集 1.7—3.5*kb* 之间的片段，经过将分离的 *DNA*

克隆进酵母载体中来制备 *cDNA* 文库 。在 *ABTS* 上从该文库中筛选出 3000 转化子。24 小时后，只有单个菌落出现阳性。从该菌落中分离质粒并测序插入片段。部分预测的氨基酸顺序与从 *RS*22 获得的片段的序列（描述同上）相对应 。图 4*SEQ. ID. NOs* :13 和 14 分别描述了完整的核苷酸和氨基酸序列 。*RSlac*3 与 *RSlac*1 显示出 48%的同源性，与 *RSlac*2 显示出 50%同源性。*RSlac*3 与毛革盖菌漆酶基因也表现出 48%的同源性。

根据本发明，使用本文提供的信息，以上述方法或作为选择的本领域任意已知方法可获得编码 *pH* 中性或碱性的丝核菌属基因。使用一种表达载体可以活性形式表达该基因。有用的表达载体含有一种能使载体稳定地整合进宿主细胞基因组或使载体在宿主细胞中独立于宿主细胞基因组进行自主复制的成份 ，和优选一个或多个使转化的宿主细胞易于选择的表型标记。表达载体也可包括编码启动子、核糖体结合位点、翻译起始信号和可任选的一种阻遏物或各种激活物基因的控制序列 。为了使表达蛋白质得以分泌，可在该基因的编码序列之前插入编码一段信号序列的核苷酸。为了在控制序列的指导下表达 ，根据本发明处理的漆酶基因在合适的阅读框中与控制基因以可操作的方式相连。可掺入质粒载体并能指导漆酶基因转录的启动子序列包括但不仅限于原核 β— 内酰胺酶启动子（*Viua — Kamaroff*，*et al*. 1978， *Proc. Natl. Acad. Sci* ，*U. SA*. 75：3727 — 3731）和 *tac* 启动子（*DeBoer*， *et al*，1983，*Proc. Natl. Acad. Sci. U. S. A*. 80：21 — 25）。进一步的资料可见“*Useful proteins from vecombinant bacteria* ”*in Scientific American* ，1980 ，242 ：74 — 94 ；和见 *Sambrook et al.* ，*Molecular Cloning*. 1989。

携带本发明的 *DNA* 构建体的表达载体可以是常规用于重组 *DNA* 方法的任意载体，载体的选择一般取决于用于导入载体的宿主细胞。因此，载体可以是一种自主复制的载体，即作为染色体外突体存在的载体，其复制不依赖于染色体复制，例如一种质粒、一种染色体外成份、极微染色体或一种人工染色体。作为选择、载体可以是一种导入宿主细胞时整合进宿主细胞基因组并与其整合进去的染色体一起复制的载体。

在载体中，*DNA* 序列应以可操作的方式与合适的启动子序列相连。启动子可以是在选定的宿主细胞中表现出转录活性的任意 *DNA* 序列并可来自编码与宿主细胞同源或异源的蛋白质的基因。指导本发明的 *DNA* 构建体转录的合适的启动子例子，特别是在细菌宿主中有大肠杆菌 *lac* 操纵子的启动子，天蓝色链霉菌琼脂酶基因 *dagA* 启动子，地衣形芽孢杆菌 α 淀粉酶基因（*amyL*）启动子，嗜热脂肪芽孢杆菌麦芽淀粉酶基因（*amyM*）启动子，解淀粉芽孢杆菌 α—淀粉酶（*amyQ*）启动子或枯草芽孢杆菌 *xylA* 和 *xylB* 基因启动子。在酵母宿主中，有用的启动子是 *eno*—1 启动子。为了在真菌宿主中转录，有用的启动子的例子是那些来自编码米曲霉高峰氏淀粉酶. *Rhizomucor miehei* 天冬氨酸蛋白酶，黑色曲霉中性 α—淀粉酶，黑色曲霉酸性稳定 α—淀粉酶，黑色曲霉或 *A*，*awamsii* 葡糖淀粉酶（*gluA*），*Rhizomucor miehei* 脂酶，米曲霉碱性蛋白霉，米曲霉磷酸丙糖异构酶或构巢曲霉乙酰胺酶的基因的启动子。优选高峰氏淀粉酶和 *gluA* 启动子。

本发明的表达载体也可包含合适的转录终止子和在真核生物中与编码本发明的漆酶的 *DNA* 序列以可操纵的方式相连的多聚

腺苷酸化序列。终止和多聚腺苷酸化序列可适当源于启动子相同的来源。载体可进一步包含使载体在其宿主细胞中复制的 *DNA* 序列。这类序列的例子是质粒 *pUC*19 ，*pACYC*177，*pUB*110，*pE*194，*pAMB*1 和 *pIJ*702 的复制起点。

载体也可包含可选择的标记，例如，其产物弥补宿主细胞缺陷的基因，如枯草芽孢杆菌或地衣形芽孢杆菌的 *dal* 基因，或提供抗生素抗性(如氨苄青霉素、卡那霉素，氯霉素或四环素的基因。曲霉属选择标记的例子包括 *amdS* ，*pyrG*，*argB*，*niaD* 和 *sC* 和一种产生潮霉素抗性的标记。优选用于曲霉属宿主细胞的是构巢曲霉或米曲霉 *amdS* 和 *pyrG* 标记。常用的哺乳动物标记是二氢叶酸还原酶(*DHFR*)基因。而且可经共转化完成选择，例如，如 *WO* 91/17243 所述。

一般优选细胞外表达。因此，本发明的漆酶可包含允许将表达蛋白质分泌进培养基的前区肽。如果需要，该前区肽可以是本发明漆酶的天然存在或常规经编码各自前区肽的 *DNA* 序列取代完成的用不不同前区肽或信号序列的取代。例如，前区可以来自曲霉属种类的葡糖淀粉酶或淀粉酶基因、芽孢杆菌属种类的淀粉酶基因、*Rhizomucor miehei* 的脂酶或蛋白酶基因、啤酒糖酵母的 α—因子基因成小牛凝乳酶原基因。特别优选的是，当宿主是真菌细胞时，前区是米曲霉高峰氏淀粉酶、黑色曲霉中性淀粉酶、芽孔杆菌 *NCIB* 11837 的麦芽淀粉酶，嗜热脂肪芽孢杆菌 α—淀粉酶，或地衣形芽孢杆菌枯草杆菌蛋白酶的前区。有效的信号序列是米曲霉高峰氏淀粉酶信号，*Rhizomucor miehei* 天冬氨酸蛋白酶信号和 *Rhizomucor miehei* 脂酶信号。

用于分别连接本发明的*DNA*构建体、启动子、终止子或其它成份以及用于将它们插入含有复制所必需的信息的合适载体中的方法是本领域的技术人员所熟知的（参见，例如：*Sambrook etal. Molecular Cloning*，1989）。

包含上面限定的本发明的*DNA*构建体或表达载体的本发明的细胞可优先应用作本发明的酶的重组生产中的宿主细胞。可用本发明的*DNA*构建体转化该细胞，通常经过将*DNA*构建体整合进宿主染色体中。一般认为这种整合对于*DNA*序列可能更稳定地保留于细胞中是一种优点。可根据常规方法实现将*DNA*构建体整合进宿主染色体中，例如经过同源或异源重组。作为选择，可用上面所述的与不同类型宿主细胞相连系的表达载体来转化细胞。

宿主细胞可选自原核细胞，如细菌细胞。合适的细菌例子是革兰氏阳性细菌、如枯草芽孢杆菌、地衣形芽孢杆菌、缓慢芽孢杆菌、短芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、解淀粉芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、*BaciUus lautus*、巨大芽孢杆菌，苏云金芽孢杆菌、或变青链霉菌或鼠灰链霉菌，或革兰氏阴性细菌，如大肠杆菌。细菌的转化可用诸如原生质体转化或用本身已知方式的感受态细胞来实现。

宿主细胞也可以是真核生物，如哺乳动物细胞、昆虫细胞、植物细胞或优选真菌细胞（包括酵母和丝状真菌）。例如，有用的哺乳动物细胞包括*CHO*或*COS*细胞。酵母宿主细胞可选自糖酵母属或裂殖糖酵母属的种类，例如啤酒酵母。有用的丝状真菌可选自曲霉属种类，例如米曲霉或黑色曲霉。作为选择，镰孢属种类的菌株（例如尖孢镰孢）可用作宿主细胞。真菌细胞的转化方法包含原生质

体形成和原生质体的转化及随后以本身已知的方式再生细胞壁。转化曲霉属宿主细胞的合适的方法在 *EP* 238023 中进行了描述。*Malardier* 等 1989 年描述了转化镰孢属种类的合适的方法。

因此，本发明提供了一种生产本发明的重组漆酶的方法，该方法包含在诱导酶产生的条件下培养上述宿主细胞并从细胞和/或培养基中回收该酶。用于培养该细胞的培养基可以是任何适于生长该宿主细胞并获得本发明的漆酶表达的常规培养基。合适的培养基可从供应商买到或根据出版的配方制备(例如，见美国标准菌种保藏所的目录。

可用常规方法从培养基中回收所得的酶，包括以离心或过滤从培养基中分离细胞，以盐法(例如硫酸铵)沉淀上清或滤液中的蛋白质成份，接着以各种色谱法纯化(例如离子交换色谱、凝胶过滤色谱、亲和色谱或类似方法)。优选的是分离的蛋白质经 *SDS*—*PAGE* 测定的纯度为大约 90%，其纯度在食品，汁液或去污剂应用中是极为重要的。

在特别优选的实施例中，漆酶的表达是在真菌宿主细胞中实现的，如在曲霉属中。如下面实施例的详细描述，漆酶基因连接到含米曲霉高峰氏 α—淀粉酶启动子和构巢曲霉 *amdS* 可选择标记的质粒上。作为选择，*amdS* 可在单独的质粒上并用于共转化。根据 *Velton* 等(*PNAS USA* 81 :1470—1474，1984)描述的方法，可将质粒(或多个质粒)用于转化曲霉属种类的宿主细胞(如米曲霉或黑色曲霉)。

本领域的技术人员将认识到本发明并不局限于使用本文特别公开的核酸片段(例如图 1 和 2)。显而易见的是本发明还包含编码

与图 1,2 和 3 所 示相同的氨基酸序列的核苷酸序列，但由于遗传密码的简并性，它们与那些特别描述的核苷酸序列不同。另外，本发明还包含其它的核苷酸片段及由其编码的蛋白质，该核苷酸片段编码与茄丝核菌漆酶蛋白质具有基本相同的最适 *pH* 并与茄丝核菌氨基酸序列表现出明显的同源水平的漆酶蛋白质。例如，现有资料表明不只一种丝核菌属的种类产生具有所述 *pH* 特征的漆酶；因此期望其它的丝核菌属种类也产生相似的漆酶，从而使用本文描述的技术可将它们用作本发明范围内的基因的来源。正如提供的实施例所示，不仅不只一种编码具有所需特征的漆酶的核苷酸和氨基酸序列，而且还有大量在序列耐受范围内仍产生功能性酶的变异体。因此，本发明还包含任何变异的核苷酸序列及由其编码的蛋白质，该蛋白质与图 1,2 和 3 描述的一个或另一个氨基酸序列相比至少保留了大约 80% 的同源性，并且保留了本文所述序列的漆酶和最适 *pH* 活性。特别是在丝核菌属漆酶高保守区保留高水平(即≥80%)同源性的变异体是可以预料到的。这类区域鉴定为 ***RSLAC***1 中的残基 458—469，***RSLAC***2 中的 478—489；***RSLACI*** 中的 131—144 和 ***RSLAC***2 中的 132—145。

例如，上面限定的种类中有用的变异体包括发生了保守氨基酸置换的变异体，其中的置换并不显著影响到蛋白质的活性。保守置换是指相同类的氨基酸可被该类型的任何其它氨基酸取代。例如，非极性脂肪族残基 *Ala*，*Val*，*Leu* 和 *Ile* 可互相交换，碱性残基 *Lys* 和 *Arg* 或酸性残基 *Asp* 和 *Glu* 也一样。同样地，*Ser* 和 *Thr* 互相是保守置换，*Asn* 和 *Gln* 也一样。这种置换可在对分子功能较重要的区域之外进行且仍产生活性酶，这对本领域的技术人员而言

是显而易见的 。在 0.1*M* 磷酸钠（*pH*7.0）中使用标准 *ABTS* 氧化方法易于测定所需活性的保留。

该蛋白质可用于许多不同的工业生产中 ；仅管该酶在低 *pH* 下也发挥一定程度的功能 ，但茄丝核菌漆酶最有利的是用于通常在中性或碱性 *pH* 下进行的生产中，因为其它漆酶在这一 *pH* 范围没有活性。这些生产包括木质素（牛皮纸和木质素硫酸盐）在溶液中的聚合以产生具有高分子量的木质素 。中性/碱性漆酶对于在较高 *pH* 下更易溶的牛皮纸木质素具有特别优点。例如，这类方法在 ***Jin et al***，***Holzforschung*** 45(6)：467－468，1991；美国专利号 4，432，921；***EP***0275544；***PCT/DK***93/00217，1992 中普进行了描述。

本发明的漆酶也可用于牛皮纸浆中木质素的原位解聚 ，由此能产生具有低木质数含量的纸浆。通常使用氯气用于木质素的解聚，它导致产生氯化的芳香族化合物，该化合物是一种对环境有害的造纸厂副产品 ，漆酶的使用改进了这一现状 。例如 ，这类用途在 ***Current opinion in Biotechnology*** 3 ：261－266 ，1992；***J. Biotechnol.*** 25：333－339，1992；***Hiroi et al***，***Svensk papperstidning*** 5：162－166，1976 中进行了描述。由于造纸厂的环境一般是碱性，在这一用途中，该漆酶比其它在酸性条件下发挥最佳功能的已知漆酶相比更有用。

染料和其它有色化合物的氧化导致化合物脱色 。漆酶 可用于该目的，它在织物间不需染料转移的情况下（例如在纺织工业和去污工业中）特别具有优势 。染料转移抑制和染料氧化的方法可见 ***WO*** 92/01406，***WO*** 92/18683 ，***EP*** 0495836 和 ***Calvo. Mededelingen van de Faculteit Landbouw － wetenschappen/Rijiksuniversitet***

Gent. 56:1565—1567,1991。

该漆酶也可用于存在于液体中的酚化合物的聚合。这种用途的一个例子是汁液的处理（例如苹果汁），以使漆酶加速存在于汁液中的酚化合物的沉淀，从而产生更稳定的汁液。这类应用在 *Stutz*, *Fruit processing* 7193,248—252,1993; *Maier et al.*, *Dt*, *Lebensmittel—rindschall* 86 (5):137—142,1990; *Dietrich et al.*, *Fluss*, *Obst* 57(2):67—73,1990 中进行了描述。以下面非限制性的实施例进一步解释本发明。

实施例

1. 茄丝核菌漆酶的纯化和鉴定

经过将含活生长物的一小块琼脂转移到 500*ml* 摇瓶中的 100*ml* *cFM*（24.0*g* 马铃薯葡萄糖培养基，3.0*g* 酵母提取物，1.0*ml* 微量元素溶液〔0.80*g* KH_2PO_4，0.64*g* $CuSO_4.5H_2O$，0.11*g* $FeSO_4.7H_2O$，0.80*g* $MnCl_2.4H_2O$，0.15*g* $ZnSO_4.7H_2O$，加蒸馏水至 1000*ml*〕加蒸馏水至 1000*ml*）中，检查在马铃薯葡萄糖琼脂（*Difco*）上培养的茄丝核菌单个分离物的漆酶形成。培养在室温下、有轨摇床上以 200*rpm* 进行。

在 50、74、122 和 170 小时时收获样品，离心并用 *ABTS*（*ABTS* =2,2'—二连氮（3 乙基苯并噻唑啉—6—硫酸）分析清亮的上清液中的漆酶。经过加入 200*μl* 2*mM* *ABTS* 的 0.1*M* 磷酸盐缓冲液（*pH*7）室温（大约 23—25℃）培养 30 分钟后在 418*mm* 处观察吸光率变化来进行分析。该方法是对 *Putter* 和 *Becker*（*Peroxidases*, *in*: *Bergmeyer*, *H.U. Ced*）, *Methods of Enzymatic Analysis*, 3*rd ed.*, *VolIII*, *pp.* 286—293,1983）的过氧化物酶分析法的改进。

电泳分离每份在172小时收获的漆酶并以*ABTS*作为色素原染色。出现一些明显的带型；菌株*RS*22显示出产生具有碱性*pI*的漆酶，将它选来作进一步鉴定。

使用丁香醛连氮方法也可测定漆酶活性。在需氧条件下，漆酶催化丁香醛连氮氧化成四甲氧基偶氮双甲又醌，颜色曲黄变柴。在1*cm*比色杯中将3000*μl* 25*mM*乙酸缓冲液（含10*mg/l*硫酸铜，5H_2O(*pH*5.5）在30℃下与225*μl* 0.28*mM*丁香醛连氮(5*mg*(溶于25*ml*乙醇中)用去矿质水调至50*ml*)混合。然后将混合物与100*μl*漆酶稀释液〔在乙酸缓冲液中稀释，以便吸光率的增加（△*OD*)在0.1—0.6范围内〕混合。反应混合物放于30℃恒温分光光度计中，从加入漆酶起反应在530*nm*下进行10到70秒。酶活性计算式为：△*OD*/分*X*0.677*X*稀释系数，以*LACU*表达。

为了纯化丝核菌属漆酶，将2.1升*LACU*活性为0.19*LACU*/*ml*的培养基通过10*μm*滤膜过滤并使用临界值为10*kDa*的*Filtron minisette OMEGA*膜经超滤浓缩至230*ml*。样品的*pH*是5.3，浓缩样品的活性测定为3.34*LACU*/*ml*。

将*pH*调至4.5后，由于有轻微沉淀，所以进行了过滤。将样品加到用20*mM*乙酸缓冲液（*pH*4.5，缓冲液*A*)平衡的40*ml S Sepharose Fast Flow*柱上。在缓冲液*A*中洗涤柱并用含1*M NaCl*的缓冲液*A*洗脱。收集活性馏分并合并。使用装配有*Diaflo YM*10膜的*Amicon*超滤装置浓缩活性合并物并将缓冲液变成缓冲液*A*。除了用超过30倍柱体积的从缓冲液*A*到含0.5*M NaCl*的缓冲液*A*的线型梯度溶液进行洗脱外，使用上述方法在5*ml S Sepharose High Performance*柱上对该样品进行再次色谱。合并这次纯化中

表现出最高活性的馏分。以A280*nm* 下一个吸光率单位等于1*mg/ml* 估测蛋白质浓度时，获得大约45*mg* 漆酶。以*SDS—PAGE* 判定蛋白质纯度＞90%。以*SDS—PAGE* 估测的分子量是大约67*KDa*。纯化蛋白质的比活是1 *LACU/mg*。使用丁香醛连氮作底物得到的纯化蛋白质的*pH* 适应状况如下面表1所示。

表1：

pH	5	6	7	8
%活性	0.5	31	100	59

为了测序蛋白质，使用无色杆菌属赖氨酸特异性蛋白质(无色杆菌属蛋白酶*I*)或地衣形芽孢杆菌谷氨酸特异性蛋白酶产生多肽。使用含0.08%*TFA* 水溶液的80% 2—丙醇（溶剂*B*)在0.1% *TFA* 水溶液(溶剂*A*)中的线型梯度以反相*HPLC* 纯化多肽。

根据厂商说明书在*Applied Biosystems* 473*A* 蛋白质测序仪上对完整蛋白质和纯化的多肽进行*N*—末端氨基酸序列分析。分离的蛋白质的起始部分的测序得到下面*N*—末端序列：

AVRNYKFDIKNVNVAPDGFQRPIVSV(*SEQ. ID. NO*:5)

然后用赖氨酸或谷氨酸特异性蛋白酶消化该蛋白质，并鉴定了下面的另一些多肽。多肽1—4可与毛革盖菌漆酶序列进行序列对比：

多肽1:*SQYVDGLRGPLVIYDPDDDH*(*SEQ. ID. NO*:6)

多肽2:*GLALVFAEAPSQIRQGVQSVQPDDA*(*SEQ. ID. NO*:7)

多肽3:*SRYBVBBASTVVMLEBWYHTPAXVLE*(*SEQ. ID. NO*:8)

多肽4:*SLGPTPNYVNPXIRDVVRVGGTTVV*(*SEQ. ID. NO*:9)

多肽 5:*IRYVGGPAVX(N)RSVI(SEQ. ID. NO*:10)

多肽 6:*ILANPA(SEQ. ID. NO*:11)

多肽 7:*YEAPSLPT(SEQ. ID. NO*:12)

在上面序列中，*X* 指未鉴定的残基，*B* 代表残基是天冬氨酸或天冬酰胺。

2. 茄丝核菌漆酶基因的分离

对从非丝核菌属种类获得的真菌性漆酶的已知氨基酸序列(*Choi et al.*. 出处同上;*German et al.*，出处同上;*Saloheimo et al.* 出处同上；和 *Kojima et al*，出处同上)进行研究以确定在它们中存在的相同序列。发现在靠近该蛋白质氨基末端第三个残基处有 2 个高度一致的区域，即 *IHWHGFFQ* 帮 *TFWYHSH*。根据这些相同序列和相应的 *DNA* 序列，使用 *Applide Biosystem* 394 *DNA/RNA* 合成仪合成了 3 个简并寡核苷酸，即 *O－lac2*〔*TGG/AAAGAC-CATA/GGTGTCG/AGTA/G*〕、其互补的 *O－lac2r* 和 *O－lac3*〔*ATC-CAT/CTGGCAT/cGGG/cA/TTCTTCCAG/A*〕。

合成的寡核苷酸用于聚合酶链式反应 (*PCR*)以便以茄丝核菌基因组 *DNA* 中筛选漆酶基因或基片段。为了扩增基因组 *DNA*，将 0.5*μg* 基因组 *DNA* 与每种引物 1*μM*，200*μM dNTP*，1*U Taq* 聚合酶 (*Boehringer Mannheim*) 在 10*mM Tris－Cl*，1.5*mM* $MgCl_2$，50*mM KCl*，1*mg/ml* 明胶；*pH*8.3 的溶液中温育。反应温育方式为在 95℃1*X*5 分钟，30*X*〔95℃1 分钟，50－60℃1 分钟，72℃1 分钟〕和 72℃下 1*X*5 分钟。*PCR* 反应扩增出一条 220 核苷酸的 *DNA* 片段。根据厂商说明书将 *PCR* 产物克隆进 *TA* 克隆载体 (*InVitrogen Corp.*)中。经过对各个克隆的 *DNA* 测序鉴定 *PCR* 产物，区分出 2

个单独的漆酶基因，命名为 *RSlac*1 和 *RSlac*2。

为了制备茄丝核菌基因组文库，用限制性酶 *Sau* 3*A* 部分消化茄丝核菌 *DNA*，在 *Tris*/乙酸盐/*EDTA* 缓冲液中通过 0.8%*Sea Plaque Agarose*（*FMC Bioproducts*）电泳以分离那些大小在 8.0 至 21*kb* 之间的 *DNA* 片段。根据厂商说明书用 *Beta*—*Agarase*（*New Eng Land Biolabs*）进一步纯化凝胶分离的片段并连接到具有 *BamHI* 末端的 λ 噬菌体 *EMBL*3 臂上。使用 *Gigapack II* 包装提取物（*stratagene*）体外包装所得的噬菌体。将 25*ml TB* 培养基＋0.2% 麦芽糖和 10*mM* $MgSO_4$ 的溶液接种到 50*μl E. coli k*802（*supE*，*hsdR*，*gal*，*metB*）过夜培养物的等份试样中并在 37℃下摇动培养至 *A*600＝0.5。将 25*μl* 1:10 和 1:50 稀释的包装噬菌体与 250*μl K*802 细胞合并在 37℃下培养 20 分钟。在 48℃下向每个稀释液中加入 5*μl* 融化的顶级琼脂。然后将混合物涂布到预温的 *LB* 平板上并在 37℃培养至少 12 小时。从这些噬菌体里在 *E. coli k*802 中产生了 170,000 噬菌斑的文库，扩增 100 倍备用。

为了筛选漆酶基因，使用常规方法将扩增的基因组文库的 25,000 噬菌斑涂布到 *NZY*/琼脂糖平板上以便转移噬菌斑。使用随机标记成 $5X10^8$*cpm*/*μg* 的 220 核苷酸 *PCR* 片段作探针筛选滤膜。滤膜在 42℃下在 50% 甲酰胺，6*xssc* 中杂交 16 小时，在 65℃下用 0.5*xssc*，0.1%*SDS* 洗涤。排出阳性克隆并使用常规方法再次筛选。鉴定的 9 个阳性克隆经 *DNA* 序列分析分成 2 类，显示出编码 2 种不同的漆酶基因（*RSlacl* 和 *RSlac*2）。使用荧光核苷酸和 *Applied Biosystems* 自动 *DNA* 测序仪（*Model* 363*A*，*Version* 1,2,0）确定每个基因的完整核苷酸序列。核苷酸和预测的氨基酸序列在图 1 和 2

中描述。

为了分离 *RSlac*3，从在 1*mM* 茴香胺存在的条件下生长的茄丝核菌菌丝体中纯化的 *polyA RNA* 用作使用标准方法进行 *cDNA* 合成的模板 。通过以 0.8% 琼脂糖凝胶电泳分离 *cDNA* 并收集大小在 1.7 和 3.5*kb* 之间的 *DNA* 片段然后将经过大小分离的 *cDNA* 克隆进酵母表达载体 *pYESZ* 来构建文库 。起初将该文库的 3000 酵母转化子涂布到含 2% 葡萄糖的 *YNB*（每升 1.7*g* 无氨基酸酵母氮源，5*g*$(NH_4)_2SO_4$ 上。在 30℃下生长 4 天后，所得的菌落重复涂布到含 0.1% 葡萄糖 、2% 半乳糖和 2*mM ABTS*〔2，2'—二连氮（3—乙基苯并噻唑啉—6—磺酸）；*Sigma* #*A*—1888〕的 *YNB* 上。在 30℃下生长 24 小时后，有一个菌落具有亮绿色的晕圈 ，逐渐变成暗紫色。分离该菌落的质粒并测序插入片段 。*RSlac*3 基因翻译部分和蛋白质的序列如图 4*SEQ. ID. NOs*. 13 和 14 所示。

3. 漆酶基因的表达

质粒 *pMWR*—1 是含高峰氏淀粉酶转录调控信号和高峰氏淀粉酶信号序列的载体的 *pUC* 。在信号序列裂解位点的单个 *sfiI* 位点和 3' 相邻的 *NsiI* 位点加工该质粒以便这两个限制性酶可用于在阅读框中导入外源蛋白质 。使用 *PCR* 反应（按上述进行，但以 100*ng* 合适的线型化质粒 *DNA* 作模板 ）和诱变引物，将 *sfiI* 位点分别在 *RSlac*1 和 *RSlac*2 的氨基酸 12 和氨基酸 14 位处导入 ，以便蛋白质的编码序列在具有高峰氏淀粉酶信号序列的阅读框中。另外，*PCR* 扩增也用于在 *RSlac*1 的 3' 端导入 *PstI* 位点（*cTGCAG*）和在 *RSlac*2 的 3' 端导入 *NsiI* 位点（*ATGCAT*）。

为了准备转化，将米曲霉细胞在 *YPG*（1*g*/*l* 酵母提取物，0.25*g*

K_2PO_4、0.125g（$MgSO_4$，3.75g 葡萄糖）中在 34℃下以 100—120rpm 培养 16—20 小时，然后以 miracloth 过滤收集。用 Mg 溶液洗涤细胞(0.6M $MgSO_4 \cdot 7H_2O$)，然后将 2—6g 细胞悬浮于 10ml MgP(1.2M $MgSO_4 \cdot 7H_2O$，10mM $NaH_2PO_4 \cdot 2H_2O$；pH5.8)中。向其中加入 1ml Novozyme^R 234 (120mg/ml MgP)，将样品在冰上放置 5 分钟。加入 1ml BSA(12mg/ml)，在 34—37℃下轻轻摇动样品。通过 miracloth 过滤收集原生质体，用 5ml ST(0.6M 山梨醇，100mM Tris，pH7)覆盖。样品以 2500rpm 离心 15 分钟，收集原生质体带。加入 2 体积的 STC(1.2M 山梨醇，10mM Tris，10mM $CaCl_2 \cdot 2H_2O$；pH7.5)，样品在 2500rpm 下离心 5 分钟。用 5ml STC 洗涤沉淀 2 次，将原生质体悬浮于 0.5—1ml STC 中。

为了进行转化，将原生质体浓度调到 1—5X10^7/ml。向 100μl 原生质体溶液中加入最大量为 10μl 的 DNA 溶液（5—10μg 超螺旋 DNA)和 0.2ml PEG（60% PEG 4000，10mM Tris，10mM $CaCl_2 \cdot H_2O$；pH7.5），将混合物充分混匀。样品在室温下静置 25 分钟，然后向其中先加入 0.2ml PEG 并混匀，然后加入 0.85ml PEG 并混匀。混合物在室温下静置 20 分钟，然后以 4000rpm 离心 15 分钟。以 2ml STC 经过在 2500rpm 离心 10 分钟洗涤沉淀。将原生质体重悬于 0.2—0.5ml STC 中，然后涂布到 COVE 平板上。COVE 培养基（pH7）含有 342.3g/l 蔗糖，25g/l 琼脂和盐溶液，盐溶液含有 26g/l KCl，26g/l $MgSO_4 \cdot H_2O$，76g/l KH_2PO_4 和 5ml/l 的微量金属，微量金属是 40mg/l $NaB_4O_7 \cdot 10H_2O$，400mg/l $CuSO_4 \cdot 5H_2O$，1200mg/l $FeSO_4 \cdot 7H_2O$，700mg/l $MnSO_4 \cdot H_2O$. 800mg/l$Na_2MO_2 \cdot 2H_2O$，10g/l $ZnSO_4 \cdot 7H_2O$。高压灭菌后，向溶液中加入 10ml/l

1*M* 过滤的乙酰胺和 5—10*ml* 3*M CsCl*。经过在含乙酰胺作碳源的 *COVE* 培养基上生长细胞来选择转化子。

样品中漆酶生产可经过接上 *ABTS* 氧化方法在含 2*mM ABTS*. *pH*5—7.3 的 *COVE* 培养基上予以证实。*RSlac*1 和 *RSlac*2 两者在 *pH*5 和 *pH*7 时表达出漆酶活性，而对照漆酶相反，它在 *pH*7.3 时基本上无活性。

使用丁香醛连氮作底物在各种 *pH* 下测定 *RSlac*1 和 *RSlac*2 每种表达的产物的氧化酶活性。测定基本上按照上对天然蛋白质的测定方法在 *pH*4—9 的范围内进行。如图 5 和 6 所示，两漆酶在 *pH* 超过 5 时具有活性，*RSlac*1 在 *pH* 超过 6 时具有特别好的活性 。将该活性方式与从 *RS*22 中分离的 *RSlac*3 漆酶所观察到的进行一般性比较(见上面表 1)，*RSlac*1 表现出最大范围的活性。

生物学材料的保藏

下列生物学材料按布达佩斯条约保藏在位于 *Bakeham Lane* ，*Egham*，*Surrey Twzo* 9*Ty* 的遗传资源资料保藏中心的国际微生物学研究所，并给出下列保藏号。

保藏品	保藏号
茄丝核菌 *RS*22	*IMI CC* 358730

下列生物材料按布达佩斯条约保藏在 1815*Vnirersity Street*，*Peoria* ，*IUinois*，61604 北方区域研究中心的农作物研究服务社专利培养的保藏中心，并给出下列保藏号。

保藏品	保藏号
含与 α—淀粉酶信号序列融合的 *RSlac*1 的大肠杆菌(*EMC*00844)	*NRRL B*—21141

含具有 *sfiI* 位点插入的 *RSlac*2 的大肠杆菌(*EMCC* 00845)	*NRRL B*—21142
含 *RSlac*3 的大肠杆菌(*EMCC* 0088)	*NRRL B*—21156

序列描述

(1)一般资料

(i)申请人:

(A)名称:Novo Nordisk A/S

(B)街道:Novo Alle

(C)城市:Bagsvrerd

(D)国家:丹麦

(E)邮编(编码):DK—2880

(F)电话:+45 4444 8888

(G)电传:+45 4449 3256

(F)电极:37304

(i)申请人

(A)名称:Novo Nordisk Biotech. Inc

(B)街道:1445 Drew Avenue

(D)国家:美国

(E)邮编(编码):95616—4880

(F)电话:(916)757—8100

(G)电传:(916)758—0317

(ii)发明题目:纯化的 pH 中性漆酶和编码该酶的核酸

(iii) 序列数: 14

(iv)联系地址:

(A)联系人:Novo Nordisk of North America . Inc.

(B)街道:405 Lexington avencce, Suite 6400

(C)城市:纽约

(D)州:纽约

(E)国家:美国

(F)编码:10174—6401

(V)计算机可读形式:

(A)媒体类型:Floppy 盘

(B)计算机:IBM PC 兼容性

(C)操作系统:PC—DOS/MS—DOS

(D)软件:Patent In Release #1.0, Version #1.25

(vi)最近申请资料:

(A)申请号:即将指定

(B)申请日:1994 年 9 月 13 日

(vii)优先权申请资料:

(A)申请号:US 08/172,331

(B)申请日:1993 年 12 月 22 日

(vii)优先权申请资料:

(A)申请号:US 08/122,230

(B)申请日:1993 年 9 月 17 日

(vii)优先权申请资料:

(A)申请号:US 08/122,827

(B)申请日:1993 年 9 月 17 日

(vii)优先权申请资料:

(A)申请号:US 08/162,827

(B)申请日:1993 年 12 月 3 日

(viii)代理人/代理资料

(A)姓名:Lowney Dr. ,Karen A.

(B)登记号:31,274

(C)参考/目录号:4052.204—WO

(ix)电传资料:

(A)电话:212—867—0123

(B)电话:212—878—9655

(2)SEQ ID NO:1 资料

(i)序列特征:

(A)长度:2838 碱基对

(B)类型:核酸

(C)链型:单链

(D)拓扑:线型

(ii)分子类型:DNA(基因组)

(vi)最初来源

(A)生物体:丝核菌属漆酶

(ix)特征

(A)名称/键:内含子

(B)位置:302..351

(ix)特征

(A)名称/键:内含子

(B)位置:463..512

(ix)特征

(A)名称/键:内含子

(B)位置:576..633

(ix)特征

(A)名称/键:内含子

(B)位置:760..818

(ix)特征

(A)名称/键:内含子

(B)位置:822..877

(ix)特征

(A)名称/键:内含子

(B)位置:1001..1054

(ix)特征

(A)名称/键:内含子

(B)位置:1316..1372

(ix)特征

(A)名称/键:内含子

(B)位置:1697..1754

(ix)特征

(A)名称/键:内含子

(B)位置:1827..1880

(ix)特征

(A)名称/键:内含子

(B)位置:1992..2051

(ix)特征

(A)名称/键:内含子

(B)位置:2157..2206

(ix)特征

(A)名称/键:内含子

(B)位置:2348..2404

(ix)特征:

(A)名称/键:内含子

(B)位置:2438...2498

(ix)特征

(A)名称/键:CDS

(B)位置:连接 (170..301, 352..462, 513..575, 634..759, 819..821, 878..1000, 1055..1315, 1373..1696, 1755..1826, 1881..1991, 2052..2156, 2207..2347, 2405..2437, 2499..2621)

(xi)序列描述:SEQ ID NO:1

```
AGCGTCACAC CAGACATCGG ATGAAAACGG AAAGTGTATG CGCCATTTGA CGTCTGCGGC      60

AACCACTGTT CATCTCGCGA GCTAACATGG GCGACGTATA AGAAGAACGC GAGAATGGGC     120

AGATTTCGAT ATCCCCTCTC GTCTCGGTTT TGGTCTCGGC TTGCCTCTA ATG GCG         175
                                                          Met Ala
                                                            1

CGC ACC ACT TTC CTT GTC TCG GTT TCG CTC TTT GTT TCC GCT GTT CTT        223
Arg Thr Thr Phe Leu Val Ser Val Ser Leu Phe Val Ser Ala Val Leu
          5                  10                  15

GCG CGC ACC GTC GAG TAC GGC TTG AAG ATT AGT GAT GGG GAG ATA GCT        271
Ala Arg Thr Val Glu Tyr Gly Leu Lys Ile Ser Asp Gly Glu Ile Ala
     20                  25                  30

CCT GAC GGT GTT AAG CGT AAT GCG ACT TTG GTACGCACTC CTTGTAATCC         321
Pro Asp Gly Val Lys Arg Asn Ala Thr Leu
 35                  40

AACAATTCAA GGTTTCTGAT GCTTGGTCAG GTA AAT GGA GGG TAT CCC GGT CCA      375
                                Val Asn Gly Gly Tyr Pro Gly Pro
                                 45                  50

CTC ATT TTT GCC AAC AAG GGG GAT ACT CTC AAA GTC AAG GTC CAA AAC       423
Leu Ile Phe Ala Asn Lys Gly Asp Thr Leu Lys Val Lys Val Gln Asn
          55                  60                  65

AAG CTC ACG AAT CCT GAG ATG TAT CGC ACC ACT TCC ATC GTATGTTCGT        472
Lys Leu Thr Asn Pro Glu Met Tyr Arg Thr Thr Ser Ile
     70                  75                  80

TCGATATCTA CTAATACATC CGTCGCTAAA TATCTTGTAG CAT TGG CAC GGT CTC       527
                                          His Trp His Gly Leu
                                                      85
```

```
TTA CAA CAT AGA AAC GCC GAC GAC GAC GGT CCT TCG TTC GTC ACT CAG      575
Leu Gln His Arg Asn Ala Asp Asp Asp Gly Pro Ser Phe Val Thr Gln
             90                  95                  100

GTAGGATTCT GGAAGGTTGG CCTGAACTCT CTGTTAACCG ACAACCCGAT GTCACCAG      633

TGC CCG ATT GTT CCA CGC GAG TCG TAT ACT TAC ACC ATA CCT CTG GAC      681
Cys Pro Ile Val Pro Arg Glu Ser Tyr Thr Tyr Thr Ile Pro Leu Asp
        105                 110                 115

GAT CAA ACC GGA ACC TAT TGG TAC CAT AGC CAC TTG AGT TCG CAA TAC      729
Asp Gln Thr Gly Thr Tyr Trp Tyr His Ser His Leu Ser Ser Gln Tyr
    120                 125                 130

GTT GAT GGT CTT CGA GGC CCG CTG GTA ATC GTGAGTATCT TGACTTGTCT        779
Val Asp Gly Leu Arg Gly Pro Leu Val Ile
135                 140

ACTGAAGGCA ACGAGACTAA AACAAGCGTC GATTCACAG TAT GTTCGTCTCC            831
                                          Tyr
                                          145

CCTTTATTTA GCTCTGGATC TTCATTTCTC ACGTAATACA TGATAG GAT CCC AAG       886
                                                Asp Pro Lys

GAT CCT CAC AGG CGT TTG TAT GAT GTT GAC GAT GAG AAG ACC GTC CTG      934
Asp Pro His Arg Arg Leu Tyr Asp Val Asp Asp Glu Lys Thr Val Leu
    150                 155                 160

ATC ATC GGT GAC TGG TAT CAT GAA TCG TCC AAG GCA ATC CTT GCT TCT      982
Ile Ile Gly Asp Trp Tyr His Glu Ser Ser Lys Ala Ile Leu Ala Ser
165                 170                 175                 180

GGT AAC ATT ACC CGA CAG GTAAGTGATA CATGCCGGTC CCAGAAAAAT            1030
Gly Asn Ile Thr Arg Gln
                185

TCTCTAAATT CATTTTAATT ACAG CGA CCG GTC TCT GCC ACC ATC AAC GGC       1081
                          Arg Pro Val Ser Ala Thr Ile Asn Gly
                                  190                 195

AAA GGT CGA TTT GAC CCT GAC AAC ACT CCT GCC AAC CCA GAT ACT CTG      1129
Lys Gly Arg Phe Asp Pro Asp Asn Thr Pro Ala Asn Pro Asp Thr Leu
                200                 205                 210

TAC ACC CTC AAG GTC AAG CGA GGG AAG CGC TAT CGT CTG CGT GTC ATC      1177
Tyr Thr Leu Lys Val Lys Arg Gly Lys Arg Tyr Arg Leu Arg Val Ile
            215                 220                 225

AAT AGC TCG GAG ATC GCT TCG TTC CGA TTC AGT GTG GAA GGT CAC AAG      1225
Asn Ser Ser Glu Ile Ala Ser Phe Arg Phe Ser Val Glu Gly His Lys
        230                 235                 240

GTG ACT GTG ATT GCT GCC GAT GGC GTC TCT ACC AAA CCG TAT CAG GTC      1273
Val Thr Val Ile Ala Ala Asp Gly Val Ser Thr Lys Pro Tyr Gln Val
    245                 250                 255

GAT GCG TTT GAT ATT CTA GCA GGA CAG CGC ATA GAT TGC GTC              1315
Asp Ala Phe Asp Ile Leu Ala Gly Gln Arg Ile Asp Cys Val
260                 265                 270

GTAAGTGTCG TCCGAACCCA CATCTGAGCT CAAGTGTTGA TACATGCGCG CTTATAG       1372
```

```
GTG GAG GCG AAC CAA GAA CCC GAC ACA TAC TGG ATC AAC GCA CCG CTG      1420
Val Glu Ala Asn Gln Glu Pro Asp Thr Tyr Trp Ile Asn Ala Pro Leu
    275                 280                 285

ACC AAC GTG CCC AAC AAG ACC GCT CAG GCT CTC CTC GTT TAT GAG GAG      1468
Thr Asn Val Pro Asn Lys Thr Ala Gln Ala Leu Leu Val Tyr Glu Glu
290                 295                 300                 305

GAT CGT CGG CCG TAC CAC CCT CCA AAG GGC CCG TAT CGC AAG TGG AGC      1516
Asp Arg Arg Pro Tyr His Pro Pro Lys Gly Pro Tyr Arg Lys Trp Ser
                310                 315                 320

GTC TCT GAG GCG ATC ATC AAG TAC TGG AAT CAC AAG CAC AAG CAC GGA      1564
Val Ser Glu Ala Ile Ile Lys Tyr Trp Asn His Lys His Lys His Gly
            325                 330                 335

CGT GGT TTG CTG TCT GGA CAT GGA GGT CTC AAG GCT CGG ATG ATC GAG      1612
Arg Gly Leu Leu Ser Gly His Gly Gly Leu Lys Ala Arg Met Ile Glu
        340                 345                 350

GGT AGC CAT CAT CTG CAT TCG CGC AGC GTC GTT AAG CGC CAG AAT GAG      1660

  Gly Ser His His Leu His Ser Arg Ser Val Val Lys Arg Gln Asn Glu
      355                 360                 365

  ACC ACC ACT GTT GTA ATG GAC GAG AGC AAG CTC GTT GTAAGTACCA        1706
  Thr Thr Thr Val Val Met Asp Glu Ser Lys Leu Val
  370                 375                 380

  TATTTAAAAG TTGGTTGGGT TTCGAATACT TATTTCAACT TTTCTTAG CCA CTG GAA  1763
                                                  Pro Leu Glu


  TAC CCC GGC GCT GCA TGC GGG TCT AAA CCT GCT GAC CTC GTC TTG GAT   1811
  Tyr Pro Gly Ala Ala Cys Gly Ser Lys Pro Ala Asp Leu Val Leu Asp
  385                 390                 395                 400

  CTC ACT TTT GGT TTG GTATGTAGCC AAATCGCCCA TATACAGGAT ACTGAATATT  1866
  Leu Thr Phe Gly Leu
                  405

  GTTTGTGCGT GTAG AAC TTT GCT ACC GGG CAC TGG ATG ATC AAC GGT ATC   1916
                 Asn Phe Ala Thr Gly His Trp Met Ile Asn Gly Ile
                                 410                 415

  CCA TAC GAG TCT CCC AAA ATC CCC ACA TTG CTC AAG ATC CTC ACT GAT   1964
  Pro Tyr Glu Ser Pro Lys Ile Pro Thr Leu Leu Lys Ile Leu Thr Asp
          420                 425                 430

  GAG GAC GGG GTT ACC GAG TCT GAC TTC GTATGTTCCC TTTTCGGTAT        2011
  Glu Asp Gly Val Thr Glu Ser Asp Phe
      435                 440

  CTTCGTATGC GTGCACTGAC TCGTGCTGGT GGGAATTTAG ACC AAG GAG GAG CAC   2066
                                             Thr Lys Glu Glu His
                                                     445

  ACA GTC ATA CTC CCG AAG AAC AAA TGC ATC GAA TTC AAC ATC AAG GGG   2114
  Thr Val Ile Leu Pro Lys Asn Lys Cys Ile Glu Phe Asn Ile Lys Gly
          450                 455                 460

  AAC TCG GGT ATT CCC ATT ACG CAC CCC GTA CAT CTT CAC GGT           2156
  Asn Ser Gly Ile Pro Ile Thr His Pro Val His Leu His Gly
      465                 470                 475
```

```
GTAAGTGCAT ATCGGATGGT TTACGATACT AAGGCTCATC AACTTTTTAG CAC ACT      2212
                                                          His Thr

TGG GAT GTC GTA CAA TTT GGC AAC AAC CCA CCC AAT TAT GTC AAT CCT      2260
Trp Asp Val Val Gln Phe Gly Asn Asn Pro Pro Asn Tyr Val Asn Pro
480                 485                 490                 495

CCC CGT AGG GAC GTG GTT GGC TCT ACA GAT GCG GGT GTG AGG ATT CAG      2308
Pro Arg Arg Asp Val Val Gly Ser Thr Asp Ala Gly Val Arg Ile Gln
                500                 505                 510

TTC AAG ACC GAC AAT CCA GGA CCG TGG TTC CTG CAC TGC GTGCGTCGGT       2357
Phe Lys Thr Asp Asn Pro Gly Pro Trp Phe Leu His Cys
            515                 520

CCCCATCGTC CGTTATGGTT TTTCTAATAC GTCCCATTCT ATTTTAG CAT ATT GAC      2413
                                                  His Ile Asp
                                                  525

TGG CAT CTT GAG GAG GGT TTC GCA GTGAGTACTG AGACCTAAGT GCTACTCGGC     2467
Trp His Leu Glu Glu Gly Phe Ala
        530                 535

TCATTACTGA TTACCGCATG TATGCGTCTA G ATG GTG TTT GCT GAA GCG CCC       2519
                                   Met Val Phe Ala Glu Ala Pro
                                                  540

GAA GCC GTC AAG GGA GGT CCA AAG AGC GTG GCC GTG GAC TCT CAG TGG      2567
Glu Ala Val Lys Gly Gly Pro Lys Ser Val Ala Val Asp Ser Gln Trp
        545                 550                 555

GAA GGG CTG TGT GGC AAG TAC GAC AAC TGG CTA AAA TCA AAT CCG GGC      2615
Glu Gly Leu Cys Gly Lys Tyr Asp Asn Trp Leu Lys Ser Asn Pro Gly
    560                 565                 570

CAG CTG TAGGCGTATC GCAGCCACAT TGGTGATGAT TGAAAGTTGC ATCTTGTTCC      2671
Gln Leu
575

TATAACCGGC TCTTATATAC GGGTGTCTCC CAGTAAAGTC GTAGCCCAAT TTCAGCCGAG   2731

ACAGATATTT AGTGGACTCT TACTCTTGTG TCCCATTGAC GCACATCGTT GCATCAAACC   2791

TGCTTTTTAT CGTCCCTCTT TGTAATTTGT GTTGCTGTAA TGTATCG                 2838
```

(2)SEQ ID NO:資料

(i)序列特征:

(A)长度:576 氨基酸

(D)拓扑:线型

(ii)分子类型:蛋白质

(xi)序列描述:SEQ ID NO:2

```
Met Ala Arg Thr Thr Phe Leu Val Ser Val Ser Leu Phe Val Ser Ala
 1               5                  10                  15

Val Leu Ala Arg Thr Val Glu Tyr Gly Leu Lys Ile Ser Asp Gly Glu
            20                  25                  30

Ile Ala Pro Asp Gly Val Lys Arg Asn Ala Thr Leu Val Asn Gly Gly
        35                  40                  45

Tyr Pro Gly Pro Leu Ile Phe Ala Asn Lys Gly Asp Thr Leu Lys Val
    50                  55                  60

Lys Val Gln Asn Lys Leu Thr Asn Pro Glu Met Tyr Arg Thr Thr Ser
65                  70                  75                  80

Ile His Trp His Gly Leu Leu Gln His Arg Asn Ala Asp Asp Asp Gly
                85                  90                  95

Pro Ser Phe Val Thr Gln Cys Pro Ile Val Pro Arg Glu Ser Tyr Thr
            100                 105                 110

Tyr Thr Ile Pro Leu Asp Asp Gln Thr Gly Thr Tyr Trp Tyr His Ser
        115                 120                 125

His Leu Ser Ser Gln Tyr Val Asp Gly Leu Arg Gly Pro Leu Val Ile
    130                 135                 140

Tyr Asp Pro Lys Asp Pro His Arg Arg Leu Tyr Asp Val Asp Asp Glu
145                 150                 155                 160

Lys Thr Val Leu Ile Ile Gly Asp Trp Tyr His Glu Ser Ser Lys Ala
                165                 170                 175

Ile Leu Ala Ser Gly Asn Ile Thr Arg Gln Arg Pro Val Ser Ala Thr
            180                 185                 190

Ile Asn Gly Lys Gly Arg Phe Asp Pro Asp Asn Thr Pro Ala Asn Pro
        195                 200                 205

Asp Thr Leu Tyr Thr Leu Lys Val Lys Arg Gly Lys Arg Tyr Arg Leu
    210                 215                 220

Arg Val Ile Asn Ser Ser Glu Ile Ala Ser Phe Arg Phe Ser Val Glu
225                 230                 235                 240

Gly His Lys Val Thr Val Ile Ala Ala Asp Gly Val Ser Thr Lys Pro
                245                 250                 255

Tyr Gln Val Asp Ala Phe Asp Ile Leu Ala Gly Gln Arg Ile Asp Cys
            260                 265                 270

Val Val Glu Ala Asn Gln Glu Pro Asp Thr Tyr Trp Ile Asn Ala Pro
        275                 280                 285

Leu Thr Asn Val Pro Asn Lys Thr Ala Gln Ala Leu Leu Val Tyr Glu
    290                 295                 300

Glu Asp Arg Arg Pro Tyr His Pro Pro Lys Gly Pro Tyr Arg Lys Trp
305                 310                 315                 320
```

Ser Val Ser Glu Ala Ile Ile Lys Tyr Trp Asn His Lys His Lys His
325 330 335

Gly Arg Gly Leu Leu Ser Gly His Gly Gly Leu Lys Ala Arg Met Ile
340 345 350

Glu Gly Ser His His Leu His Ser Arg Ser Val Val Lys Arg Gln Asn
355 360 365

Glu Thr Thr Thr Val Val Met Asp Glu Ser Lys Leu Val Pro Leu Glu
370 375 380

Tyr Pro Gly Ala Ala Cys Gly Ser Lys Pro Ala Asp Leu Val Leu Asp
385 390 395 400

Leu Thr Phe Gly Leu Asn Phe Ala Thr Gly His Trp Met Ile Asn Gly
405 410 415

Ile Pro Tyr Glu Ser Pro Lys Ile Pro Thr Leu Leu Lys Ile Leu Thr
420 425 430

Asp Glu Asp Gly Val Thr Glu Ser Asp Phe Thr Lys Glu Glu His Thr
435 440 445

Val Ile Leu Pro Lys Asn Lys Cys Ile Glu Phe Asn Ile Lys Gly Asn
450 455 460

Ser Gly Ile Pro Ile Thr His Pro Val His Leu His Gly His Thr Trp
465 470 475 480

Asp Val Val Gln Phe Gly Asn Asn Pro Pro Asn Tyr Val Asn Pro Pro
485 490 495

Arg Arg Asp Val Val Gly Ser Thr Asp Ala Gly Val Arg Ile Gln Phe
500 505 510

Lys Thr Asp Asn Pro Gly Pro Trp Phe Leu His Cys His Ile Asp Trp
515 520 525

His Leu Glu Glu Gly Phe Ala Met Val Phe Ala Glu Ala Pro Glu Ala
530 535 540

Val Lys Gly Gly Pro Lys Ser Val Ala Val Asp Ser Gln Trp Glu Gly
545 550 555 560

Leu Cys Gly Lys Tyr Asp Asn Trp Leu Lys Ser Asn Pro Gly Gln Leu
565 570 575

(i)序列特征：

(A)长度：3117 碱基对

(B)类型：核酸

(C)键型：单链

(D)拓扑：线型

(ii)分子类型:DNA(基因组)

(vi)最初来源:

(A)生物体:丝核菌属漆酶

(ix)特征

(A)名称/键:CDS

(B)位置:连接 393..524, 577..687, 737..799, 860..985, 1043
..1045, 1097..1219, 1269..1538, 1601..1996, 2047
..2118, 2174..2284, 2338..2439, 2495..2635, 2693
..2725, 2786..2899)

(ix)特征

(A)名称/键:内含子

(B)位置:525..576

(ix)特征

(A)名称/键:内含子

(B)位置:688..736

(ix)特征

(A)名称/键:内含子

(B)位置:800..859

(ix)特征

(A)名称/键:内含子

(B)位置:986..1042

(ix)特征

(A)名称/键:内含子

(B)位置:1220..1268

(ix)特征

(A)名称/键:内含子

(B)位置:1539..1600

(ix)特征

(A)名称/键:内含子

(B)位置:1823..1936

(ix)特征

(A)名称/键:内含子

(B)位置:1973..2046

(ix)特征

(A)名称/键:内含子

(B)位置:2119..2173

(ix)特征

(A)名称/键:内含子

(B)位置:2285..2337

(ix)特征

(A)名称/键:内含子

(B)位置:2440..2494

(ix)特征

(A)名称/键:内含子

(B)位置:2636..2692

(ix)特征

(A)名称/键:内含子

(B)位置:1046..1096

(xi)序列描述:SEQ ID NO:3:

```
GAGTGATCCG CCAGAGTTCA GGCGGATAAG TTCCTAAATA GTCATTCGCC TATTCGTGTA      60

CCTCAGCATA CTGACGACAT ACCGCCAGAT CGCCCTCGGT TCGGGCGTGG CATACGTTCG     120

CAAGGGCACC TCACGGAGCA AACTCTAAAA AGCTTCGGCA TGGATTGCAT TTTGTATTGT     180

AAACAAGTTA CGAGAAAAAC AATAGATCAG TTTTTGCCGA ATCGGATGGC TTGAAACGGA     240

AGTACCGATG GCCGATCCGA GTCGAATGAA TTAACGCATC TGAAACGGGA CCCTGAGTCG     300

AGGCACCCGC CGGCCTTGGC CGTATAAGTC ACTTGTCGCC AACTAGCACT TTTTCATTCC     360

CCCTTTTCTT CTTCCTCGTC TTCTTCTTCT CT ATG GCT CGG TCG ACT ACT TCA       413
                                      Met Ala Arg Ser Thr Thr Ser
                                        1               5

CTC TTT GCA CTG TCT CTC GTT GCT TCA GCG TTT GCT CGA GTC GTT GAC       461
Leu Phe Ala Leu Ser Leu Val Ala Ser Ala Phe Ala Arg Val Val Asp
        10                  15                  20

TAT GGG TTT GAT GTG GCT AAT GGG GCA GTT GCT CCG GAT GGT GTA ACA       509
Tyr Gly Phe Asp Val Ala Asn Gly Ala Val Ala Pro Asp Gly Val Thr
    25                  30                  35

AGG AAC GCG GTT CTC GTGAGTTAGC TGTAAGATGG TGTATATGCT GGTTGCCTAA       564
Arg Asn Ala Val Leu
 40

CGGGAATGTC AG GTC AAT GGT CGC TTC CCT GGT CCA TTG ATC ACC GCC         612
             Val Asn Gly Arg Phe Pro Gly Pro Leu Ile Thr Ala
              45                  50                  55

AAC AAG GGG GAT ACA CTT AAA ATC ACC GTG CGG AAT AAA CTC TCC GAT       660
Asn Lys Gly Asp Thr Leu Lys Ile Thr Val Arg Asn Lys Leu Ser Asp
            60                  65                  70

CCA ACT ATG CGA AGG AGC ACG ACC ATC GTTAGTACTT CCCCTCATCT             707
Pro Thr Met Arg Arg Ser Thr Thr Ile
        75                  80

GTCTTGAAAC TTTCTCATCT TTTTTGAAG CAC TGG CAC GGT CTG CTC CAA CAC       760
                                  His Trp His Gly Leu Leu Gln His
                                                  85

AGG ACG GCA GAA GAA GAT GGC CCG GCC TTT GTA ACC CAG GTATGCCTTA       809
Arg Thr Ala Glu Glu Asp Gly Pro Ala Phe Val Thr Gln
 90                  95                  100

TCCTATCGCT GCTCTGTCCC CGCGTCCTTC CCTGACTCGG GCGATTCTAG TGC CCG       865
                                                      Cys Pro
```

```
ATT CCT CCG CAA GAA TCG TAC ACC TAT ACG ATG CCG CTC GGC GAA CAG      913
Ile Pro Pro Gln Glu Ser Tyr Thr Tyr Thr Met Pro Leu Gly Glu Gln
105                 110                 115                 120

ACC GGC ACG TAT TGG TAC CAC AGC CAC TTG AGC TCC CAG TAT GTG GAC      961
Thr Gly Thr Tyr Trp Tyr His Ser His Leu Ser Ser Gln Tyr Val Asp
                125                 130                 135

GGG TTG CGT GGG CCC ATC GTT ATT GTAAGTCTTC ATTTAACCTT ATTCTTGGTT     1015
Gly Leu Arg Gly Pro Ile Val Ile
            140

ATGGCTGATT GTGACGTCGT GGTTAGT ATG TTCGTGGCTT CCACAAGAAG              1065
                              Met
                              145

TCAGCAGCCC TTGAAGCTAA CTTTATTCCA G GAC CCC CAC GAC CCG TAC AGA       1117
                                  Asp Pro His Asp Pro Tyr Arg
                                                  150

AAC TAC TAT GAT GTC GAC GAC GAG CGT ACG GTC TTT ACT TTA GCA GAC      1165
Asn Tyr Tyr Asp Val Asp Asp Glu Arg Thr Val Phe Thr Leu Ala Asp
        155                 160                 165

TGG TAC CAC ACG CCG TCG GAG GCT ATC ATT GCC ACC CAC GAT GTC TTG      1213
Trp Tyr His Thr Pro Ser Glu Ala Ile Ile Ala Thr His Asp Val Leu
    170                 175                 180

AAA ACG GTACGCGTTA ATCCTTCTAG CTTTCTTTCC TTGGGTCACT TTCTATCAG       1268
Lys Thr
185

ATC CCC GAC TCG GGT ACG ATC AAC GGC AAA GGC AAA TAC GAT CCT GCT      1316
Ile Pro Asp Ser Gly Thr Ile Asn Gly Lys Gly Lys Tyr Asp Pro Ala
            190                 195                 200

TCG GCT AAC ACC AAC AAC ACG ACA CTC GAG AAC CTC TAC ACT CTC AAA      1364
Ser Ala Asn Thr Asn Asn Thr Thr Leu Glu Asn Leu Tyr Thr Leu Lys
        205                 210                 215

GTC AAA CGC GGC AAG CGG TAT CGC CTG AGG ATT ATC AAC GCC TCG GCC      1412
Val Lys Arg Gly Lys Arg Tyr Arg Leu Arg Ile Ile Asn Ala Ser Ala
    220                 225                 230

ATC GCT TCG TTC CGG TTC GGC GTG CAG GGC CAC AAG TGC ACG ATC ATC      1460
Ile Ala Ser Phe Arg Phe Gly Val Gln Gly His Lys Cys Thr Ile Ile
235                 240                 245                 250

GAG GCT GAT GGC GTC CTC ACC AAA CCG ATC GAG GTC GAT GCG TTT GAT      1508
Glu Ala Asp Gly Val Leu Thr Lys Pro Ile Glu Val Asp Ala Phe Asp
                255                 260                 265

ATT CTA GCA GGC CAG AGG TAT AGC TGC ATC GTAAGTCTAC CTATGCCTTG       1558
Ile Leu Ala Gly Gln Arg Tyr Ser Cys Ile
            270                 275

TTGTGGAGAT AAGAACCTGA CTGAATGTAT GCGCTCCAAT AG TTG AAG GCC GAC      1612
                                              Leu Lys Ala Asp
                                                  280

CAA GAT CCT GAT TCC TAC TGG ATA AAT GCG CCA ATC ACA AAC GTT CTC      1660
Gln Asp Pro Asp Ser Tyr Trp Ile Asn Ala Pro Ile Thr Asn Val Leu
                285                 290                 295

AAC ACC AAC GTC CAG GCA TTG CTA GTG TAT GAA GAT GAC AAG CGT CCT      1708
```

```
Asn Thr Asn Val Gln Ala Leu Leu Val Tyr Glu Asp Asp Lys Arg Pro
            300                 305                 310

ACT CAC TAC CCC TGG AAG CCG TTT TTG ACA TGG AAG ATA TCA AAT GAA      1756
Thr His Tyr Pro Trp Lys Pro Phe Leu Thr Trp Lys Ile Ser Asn Glu
        315                 320                 325

ATC ATT CAG TAC TGG CAG CAC AAG CAC GGG TCG CAC GGT CAC AAG GGA      1804
Ile Ile Gln Tyr Trp Gln His Lys His Gly Ser His Gly His Lys Gly
    330                 335                 340

AAG GGG CAT CAT CAT AAA GTC CGG GCC ATT GGA GGT GTA TCC GGG TTG      1852
Lys Gly His His His Lys Val Arg Ala Ile Gly Gly Val Ser Gly Leu
345                 350                 355                 360

AGC TCC AGG GTT AAG AGC CGG GCG AGT GAC CTA TCG AAG AAG GCT GTC      1900
Ser Ser Arg Val Lys Ser Arg Ala Ser Asp Leu Ser Lys Lys Ala Val
                365                 370                 375

GAG TTG GCT GCT GCA CTC GTT GCG GGT GAG GCC GAG TTG GAC AAG AGG      1948
Glu Leu Ala Ala Ala Leu Val Ala Gly Glu Ala Glu Leu Asp Lys Arg
            380                 385                 390

CAG AAT GAG GAT AAT TCG ACT ATT GTA TTG GAT GAG ACC AAG CTT ATT      1996
Gln Asn Glu Asp Asn Ser Thr Ile Val Leu Asp Glu Thr Lys Leu Ile
        395                 400                 405

GTAAGTCCCT TAATTTTTTT CGGTGTCACG GAAGCTAACC CGCGTAATAG CCG TTG      2052
                                                        Pro Leu
                                                            410

GTT CAA CCT GGT GCA CCG GGC GGC TCC AGA CCA GCT GAC GTC GTG GTC      2100
Val Gln Pro Gly Ala Pro Gly Gly Ser Arg Pro Ala Asp Val Val Val
                415                 420                 425

CCT CTG GAC TTT GGC CTC GTATGTGGCT TCTTGTTATT CGTCCGGAAT            2148
Pro Leu Asp Phe Gly Leu
            430

GCAAACTGAT TTGGGTGGGC TATAG AAC TTT GCC AAC GGA CTG TGG ACG ATA      2200
                           Asn Phe Ala Asn Gly Leu Trp Thr Ile
                                435                 440

AAC AAT GTC TCC TAC TCC CCT CCG GAT GTC CCT ACT CTC CTC AAG ATC      2248
Asn Asn Val Ser Tyr Ser Pro Pro Asp Val Pro Thr Leu Leu Lys Ile
            445                 450                 455

TTG ACC GAC AAA GAC AAA GTC GAC GCT TCT GAC TTC GTAGGTTCCT          2294
Leu Thr Asp Lys Asp Lys Val Asp Ala Ser Asp Phe
        460                 465

CTTCTTCTTT TCAAACTAGC TACTGACATT AAGTGAACGT CAG ACG GCC GAT GAA      2349
                                                Thr Ala Asp Glu
                                                    470

CAC ACG TAT ATT CTT CCA AAG AAC CAA GTT GTC GAG TTG CAC ATC AAG      2397
His Thr Tyr Ile Leu Pro Lys Asn Gln Val Val Glu Leu His Ile Lys
    475                 480                 485

GGA CAG GCT TTG GGA ATC GTA CAC CCC CTT CAT CTG CAT GGC              2439
Gly Gln Ala Leu Gly Ile Val His Pro Leu His Leu His Gly
490                 495                 500

GTACGTCTTT CTCACACTGT TCCAGCTCCT ATTCTCTAAC ACACTCCTGC GATAG CAT    2497
                                                            His
```

Val Ala Pro Asp Gly Val Thr Arg Asn Ala Val Leu Val Asn Gly Arg
35 40 45

Phe Pro Gly Pro Leu Ile Thr Ala Asn Lys Gly Asp Thr Leu Lys Ile
50 55 60

Thr Val Arg Asn Lys Leu Ser Asp Pro Thr Met Arg Arg Ser Thr Thr
65 70 75 80

Ile His Trp His Gly Leu Leu Gln His Arg Thr Ala Glu Glu Asp Gly
85 90 95

Pro Ala Phe Val Thr Gln Cys Pro Ile Pro Pro Gln Glu Ser Tyr Thr
100 105 110

Tyr Thr Met Pro Leu Gly Glu Gln Thr Gly Thr Tyr Trp Tyr His Ser
115 120 125

His Leu Ser Ser Gln Tyr Val Asp Gly Leu Arg Gly Pro Ile Val Ile
130 135 140

Met Asp Pro His Asp Pro Tyr Arg Asn Tyr Tyr Asp Val Asp Asp Glu
145 150 155 160

Arg Thr Val Phe Thr Leu Ala Asp Trp Tyr His Thr Pro Ser Glu Ala
165 170 175

Ile Ile Ala Thr His Asp Val Leu Lys Thr Ile Pro Asp Ser Gly Thr
180 185 190

Ile Asn Gly Lys Gly Lys Tyr Asp Pro Ala Ser Ala Asn Thr Asn Asn
195 200 205

Thr Thr Leu Glu Asn Leu Tyr Thr Leu Lys Val Lys Arg Gly Lys Arg
210 215 220

Tyr Arg Leu Arg Ile Ile Asn Ala Ser Ala Ile Ala Ser Phe Arg Phe
225 230 235 240

Gly Val Gln Gly His Lys Cys Thr Ile Ile Glu Ala Asp Gly Val Leu
245 250 255

Thr Lys Pro Ile Glu Val Asp Ala Phe Asp Ile Leu Ala Gly Gln Arg
260 265 270

Tyr Ser Cys Ile Leu Lys Ala Asp Gln Asp Pro Asp Ser Tyr Trp Ile
275 280 285

Asn Ala Pro Ile Thr Asn Val Leu Asn Thr Asn Val Gln Ala Leu Leu
290 295 300

Val Tyr Glu Asp Asp Lys Arg Pro Thr His Tyr Pro Trp Lys Pro Phe
305 310 315 320

Leu Thr Trp Lys Ile Ser Asn Glu Ile Ile Gln Tyr Trp Gln His Lys
325 330 335

His Gly Ser His Gly His Lys Gly Lys Gly His His His Lys Val Arg
340 345 350

Ala Ile Gly Gly Val Ser Gly Leu Ser Ser Arg Val Lys Ser Arg Ala
355 360 365

```
GCG TTC GAC GTC GTC CAA TTC GGC GAC AAC GCT CCA AAC TAC GTG AAC      2545
Ala Phe Asp Val Val Gln Phe Gly Asp Asn Ala Pro Asn Tyr Val Asn
505                 510                 515                 520

CCT CCG CGT AGG GAT GTA GTA GGC GTA ACT GAT GCT GGA GTC CGT ATC      2593
Pro Pro Arg Arg Asp Val Val Gly Val Thr Asp Ala Gly Val Arg Ile
                525                 530                 535

CAG TTC AGA ACC GAT AAC CCG GGC CCT TGG TTC CTC CAT TGC              2635
Gln Phe Arg Thr Asp Asn Pro Gly Pro Trp Phe Leu His Cys
            540                 545                 550

GTATGCTCTT CATCTCCCAC CGCTTGTTCT TTACTTATGG TTTACCTTGC GATTTAG       2692

CAC ATT GAT TGG CAC TTG GAA GAA GGA TTT GCT GTAAGTTATT ATTCCTATTC    2745
His Ile Asp Trp His Leu Glu Glu Gly Phe Ala
                555                 560

CGAAGCATCG GGGAGATGCT AACCAAGGGT GTGTTTTAAG ATG GTA TTC GCC GAA      2800
                                        Met Val Phe Ala Glu
                                                565

GCG CCT GAA GAT ATC AAG AAA GGC TCT CAG AGT GTC AAG CCT GAC GGA      2848
Ala Pro Glu Asp Ile Lys Lys Gly Ser Gln Ser Val Lys Pro Asp Gly
            570                 575                 580

CAA TGG AAG AAA CTA TGC GAG AAG TAT GAG AAG TTG CCT GAA GCA CTG      2896
Gln Trp Lys Lys Leu Cys Glu Lys Tyr Glu Lys Leu Pro Glu Ala Leu
        585                 590                 595

CAG TGAAGTTGCA GTTGTTTCCC ATTCGGGAAC TGGCTCACTA TTCCTTTTGC           2949
Gln

ATAATTCGGA CTTTTATTTT GGGACATTAT TGGACTATGG ACTTGTTTGT CACACCCTCG    3009

CTCACTGTGT CCCTCGTTGA GTACCTATAC TCTATTCGTA TAGTGGGAAT ATGGAATATC    3069

GGATGTAATA AATGCTCGTG CGTTTGGTGC TCGAAATGGG GTAGGACT                 3117
```

(2)SEQ ID NO:4 资料

(i)序列特征:

(A)长度:599 氨基酸

(B)类型:氨基酸

(D)拓扑:线型

(ii)分子类型:蛋白质

(xi)序列描述:SEQ ID NO:4

```
Met Ala Arg Ser Thr Thr Ser Leu Phe Ala Leu Ser Leu Val Ala Ser
  1               5                  10                  15

Ala Phe Ala Arg Val Val Asp Tyr Gly Phe Asp Val Ala Asn Gly Ala
            20                  25                  30
```

Ser Asp Leu Ser Lys Lys Ala Val Glu Leu Ala Ala Ala Leu Val Ala
370 375 380

Gly Glu Ala Glu Leu Asp Lys Arg Gln Asn Glu Asp Asn Ser Thr Ile
385 390 395 400

Val Leu Asp Glu Thr Lys Leu Ile Pro Leu Val Gln Pro Gly Ala Pro
405 410 415

Gly Gly Ser Arg Pro Ala Asp Val Val Val Pro Leu Asp Phe Gly Leu
420 425 430

Asn Phe Ala Asn Gly Leu Trp Thr Ile Asn Asn Val Ser Tyr Ser Pro
435 440 445

Pro Asp Val Pro Thr Leu Leu Lys Ile Leu Thr Asp Lys Asp Lys Val
450 455 460

Asp Ala Ser Asp Phe Thr Ala Asp Glu His Thr Tyr Ile Leu Pro Lys
465 470 475 480

Asn Gln Val Val Glu Leu His Ile Lys Gly Gln Ala Leu Gly Ile Val
485 490 495

His Pro Leu His Leu His Gly His Ala Phe Asp Val Val Gln Phe Gly
500 505 510

Asp Asn Ala Pro Asn Tyr Val Asn Pro Pro Arg Arg Asp Val Val Gly
515 520 525

Val Thr Asp Ala Gly Val Arg Ile Gln Phe Arg Thr Asp Asn Pro Gly
530 535 540

Pro Trp Phe Leu His Cys His Ile Asp Trp His Leu Glu Glu Gly Phe
545 550 555 560

Ala Met Val Phe Ala Glu Ala Pro Glu Asp Ile Lys Lys Gly Ser Gln
565 570 575

Ser Val Lys Pro Asp Gly Gln Trp Lys Lys Leu Cys Glu Lys Tyr Glu
580 585 590

Lys Leu Pro Glu Ala Leu Gln
595

(xi)序列描述:SEQ ID NO:5

Ala Val Arg Asn Tyr Lys Phe Asp Ile Lys Asn Val Asn Val Ala Pro
1 5 10 15

Asp Gly Phe Gln Arg Pro Ile Val Ser Val
20 25

(2)SEQ ID NO:6 资料

(i)序列特征:

(A)长度:20 个氨基酸

(B)类型:氨基酸

(C)键型:单链

(D)拓扑:线型

(ii)分子类型:多肽

(xi)序列描述:SEQ ID NO:6

```
Ser Gln Tyr Val Asp Gly Leu Arg Gly Pro Leu Val Ile Tyr Asp Pro
1               5                   10                  15

Asp Asp Asp His
            20
```

(2)SEQ ID NO:7 资料

(i)序列特征:

(A)长度:26 个氨基酸

(B)类型:氨基酸

(C)键型:单链

(D)拓扑:线型

(ii)分子类型:多肽

(xi)序列描述:SEQ ID NO:7

```
Ser Arg Tyr Asx Val Asx Asx Ala Ser Thr Val Val Met Leu Glu Asx
1               5                   10                  15

Trp Tyr Arg Thr Pro Ala Xaa Val Leu Glu
            20                  25
```

(2)SEQ ID NO:8 资料

(i)序列特征:

(A)长度:25 个氨基酸

(B)类型:氨基酸

(C)键型:单链

(D)拓扑:线型

(ii)分子类型:多肽

(xi)序列描述:SEQ ID NO:8

```
Ser Leu Gly Pro Thr Pro Asn Tyr Val Asn Pro Xaa Ile Arg Asp Val
1               5                  10                  15

Val Arg Val Gly Gly Thr Thr Val Val
            20                  25
```

(2)SEQ ID NO:9 资料

(i)序列特征:

(A)长度:25 个氨基酸

(B)类型:氨基酸

(C)键型:单链

(D)拓扑:线型

(ii)分子类型:多肽

(xi)序列描述:SEQ ID NO:9

```
Gly Leu Ala Leu Val Phe Ala Glu Ala Pro Ser Gln Ile Arg Gln Gly
1               5                  10                  15

Val Gln Ser Val Gln Pro Asp Asp Ala
            20                  25
```

(2)SEQ ID NO:10 资料

(i)序列特征:

(A)长度:14 个氨基酸

(B)类型:氨基酸

(C)键型:单链

(D)拓扑:线型

(ii)分子类型：多肽

(xi)序列描述：SEQ ID NO：10

```
Ile Arg Tyr Val Gly Gly Pro Ala Val Xaa Arg Ser Val Ile
1               5                   10
```

(2)SEQ ID NO：11 资料

(i)序列特征：

(A)长度：6 个氨基酸

(B)类型：氨基酸

(C)键型：单链

(D)拓扑：线型

(ii)分子类型：多肽

(xi)序列描述：SEQ ID NO：11

```
Ile Leu Ala Asn Pro Ala
1               5
```

(2)SEQ ID NO：12 资料

(i)序列特征：

(A)长度：8 个氨基酸

(B)类型：氨基酸

(C)键型：单链

(D)拓扑：线型

(ii)分子类型：多肽

(xi)序列描述：SEQ ID NO：12

```
Tyr Glu Ala Pro Ser Leu Pro Thr
1               5
```

(2)SEQ ID NO：13 资料

(i)序列特征：

(A)长度： 1912 碱基对

(B)类型：核酸

(C)键型：单链

(D)拓扑：线型

(ii)分子类型：CDNA

(vi)最初来源：

(A)生物体：丝核菌属漆酶

(ix)特征

(A)名称/键：CDS

(B)位置：85:1671

(xi)序列描述：SEQ ID NO:13

```
CTAACGCTTG GTGCCGAGCT CGGATCCACT AGTAACGCGC GCCAGTGTGC TGGAATTCGC      60

GGCCGCGTCG ACACCTCCTT CAAG ATG CTT TCT AGC ATT ACC CTC CTA CCT        111
                           Met Leu Ser Ser Ile Thr Leu Leu Pro
                             1               5

TTG CTC GCT GCG GTC TCA ACC CCC GCC TTT GCT GCC GTC CGC AAC TAT      159
Leu Leu Ala Ala Val Ser Thr Pro Ala Phe Ala Ala Val Arg Asn Tyr
 10                  15                  20                  25

AAG TTC GAC ATC AAG AAC GTC AAT GTC GCT CCC GAT GGC TTT CAG CGC      207
Lys Phe Asp Ile Lys Asn Val Asn Val Ala Pro Asp Gly Phe Gln Arg
                30                  35                  40

TCT ATC GTC TCC GTC AAC GGT TTA GTT CCT GGC ACG TTG ATC ACG GCC      255
Ser Ile Val Ser Val Asn Gly Leu Val Pro Gly Thr Leu Ile Thr Ala
            45                  50                  55

AAC AAG GGT GAC ACC TTG CGC ATT AAT GTC ACG AAT CAA CTC ACG GAC      303
Asn Lys Gly Asp Thr Leu Arg Ile Asn Val Thr Asn Gln Leu Thr Asp
        60                  65                  70

CCT AGT ATG CGT CGT GCC ACA ACG ATT CAT TGG CAT GGA TTG TTC CAA      351
Pro Ser Met Arg Arg Ala Thr Thr Ile His Trp His Gly Leu Phe Gln
    75                  80                  85
```

```
GCT ACT ACC GCC GAC GAG GAT GGC CCC GCA TTC GTC ACG CAA TGC CCT      399
Ala Thr Thr Ala Asp Glu Asp Gly Pro Ala Phe Val Thr Gln Cys Pro
 90                  95                  100                 105

ATT GCG CAA AAT TTG TCC TAT ACA TAC GAG ATC CCA TTG CGC GGC CAA      447
Ile Ala Gln Asn Leu Ser Tyr Thr Tyr Glu Ile Pro Leu Arg Gly Gln
                110                 115                 120

ACA GGA ACC ATG TGG TAT CAC GCC CAT CTT GCG AGT CAA TAT GTC GAT      495
Thr Gly Thr Met Trp Tyr His Ala His Leu Ala Ser Gln Tyr Val Asp
            125                 130                 135

GGA TTG CGA GGC CCT TTG GTC ATC TAT GAT CCA AAC GAC CCA CAC AAG      543
Gly Leu Arg Gly Pro Leu Val Ile Tyr Asp Pro Asn Asp Pro His Lys
        140                 145                 150

TCG CGC TAC GAC GTG GAT GAT GCG AGC ACA GTA GTC ATG CTT GAG GAC      591
Ser Arg Tyr Asp Val Asp Asp Ala Ser Thr Val Val Met Leu Glu Asp
    155                 160                 165

TGG TAC CAT ACT CCG GCA CCC GTT CTA GAA AAG CAA ATG TTC TCG ACT      639
Trp Tyr His Thr Pro Ala Pro Val Leu Glu Lys Gln Met Phe Ser Thr
170                 175                 180                 185

AAT AAC ACC GCT CTG CTC TCT CCT GTT CCG GAC TCG GGT CTT ATC AAT      687
Asn Asn Thr Ala Leu Leu Ser Pro Val Pro Asp Ser Gly Leu Ile Asn
                190                 195                 200

GGC AAA GGG CGC TAT GTG GGC GGT CCC GCA GTT CCC CGG TCA GTA ATC      735
Gly Lys Gly Arg Tyr Val Gly Gly Pro Ala Val Pro Arg Ser Val Ile
            205                 210                 215

AAC GTA AAA CGT GGG AAA CGA TAT CGC TTG CGC GTA ATC AAC GCT TCT      783
Asn Val Lys Arg Gly Lys Arg Tyr Arg Leu Arg Val Ile Asn Ala Ser
        220                 225                 230

GCT ATC GGG TCG TTT ACC TTT TCG ATC GAA GGA CAT AGT CTG ACT GTC      831
Ala Ile Gly Ser Phe Thr Phe Ser Ile Glu Gly His Ser Leu Thr Val
    235                 240                 245

ATT GAG GCC GAT GGG ATC CTG CAC CAG CCC TTG GCT GTT GAC AGC TTC      879
Ile Glu Ala Asp Gly Ile Leu His Gln Pro Leu Ala Val Asp Ser Phe
250                 255                 260                 265

CAG ATT TAC GCT GGA CAA CGC TAC TCT GTC ATC GTT GAA GCC AAC CAA      927
Gln Ile Tyr Ala Gly Gln Arg Tyr Ser Val Ile Val Glu Ala Asn Gln
                270                 275                 280

ACC GCC GCC AAC TAC TGG ATT CGT GCA CCA ATG ACC GTT GCA GGA GCC      975
Thr Ala Ala Asn Tyr Trp Ile Arg Ala Pro Met Thr Val Ala Gly Ala
            285                 290                 295

GGA ACC AAT GCA AAC TTG GAC CCC ACC AAT GTC TTT GCC GTA TTG CAC     1023
Gly Thr Asn Ala Asn Leu Asp Pro Thr Asn Val Phe Ala Val Leu His
        300                 305                 310

TAC GAG GGA GCG CCC AAC GCC GAA CCC ACG ACG GAA CAA GGC AGT GCT     1071
Tyr Glu Gly Ala Pro Asn Ala Glu Pro Thr Thr Glu Gln Gly Ser Ala
    315                 320                 325

ATC GGT ACT GCA CTC GTT GAA GAG AAC CTC CAT GCG CTC ATC AAC CCT     1119
Ile Gly Thr Ala Leu Val Glu Glu Asn Leu His Ala Leu Ile Asn Pro
330                 335                 340                 345

GGC GCT CCG GGC GGC TCC GCT CCC GCA GAC GTT TCC CTC AAT CTT GCA     1167
Gly Ala Pro Gly Gly Ser Ala Pro Ala Asp Val Ser Leu Asn Leu Ala
                350                 355                 360
```

```
ATT GGG CGC AGC ACA GTT GAT GGG ATT CTT AGG TTC ACA TTT AAT AAC      1215
Ile Gly Arg Ser Thr Val Asp Gly Ile Leu Arg Phe Thr Phe Asn Asn
            365                 370                 375

ATC AAG TAC GAG GCT CCT TCG TTG CCC ACG CTC TTG AAG ATT TTG GCA      1263
Ile Lys Tyr Glu Ala Pro Ser Leu Pro Thr Leu Leu Lys Ile Leu Ala
        380                 385                 390

AAC AAT GCG AGC AAT GAC GCC GAT TTC ACG CCA AAT GAG CAC ACT ATC      1311
Asn Asn Ala Ser Asn Asp Ala Asp Phe Thr Pro Asn Glu His Thr Ile
    395                 400                 405

GTA TTG CCA CAC AAT AAA GTT ATC GAG CTC AAT ATC ACC GGA GGT GCA      1359
Val Leu Pro His Asn Lys Val Ile Glu Leu Asn Ile Thr Gly Gly Ala
410                 415                 420                 425

GAC CAC CCT ATC CAT CTC CAC GGC CAT GTG TTT GAT ATC GTC AAA TCA      1407
Asp His Pro Ile His Leu His Gly His Val Phe Asp Ile Val Lys Ser
                430                 435                 440

CTC GGT GGT ACC CCG AAC TAT GTC AAC CCG CCA CGC AGG GAC GTA GTT      1455
Leu Gly Gly Thr Pro Asn Tyr Val Asn Pro Pro Arg Arg Asp Val Val
            445                 450                 455

CGT GTC GGA GGC ACC GGT GTG GTA CTC CGA TTC AAG ACC GAT AAC CCA      1503
Arg Val Gly Gly Thr Gly Val Val Leu Arg Phe Lys Thr Asp Asn Pro
        460                 465                 470

GGC CCA TGG TTT GTT CAC TGC CAC ATT GAC TGG CAC TTG GAG GCT GGG      1551
Gly Pro Trp Phe Val His Cys His Ile Asp Trp His Leu Glu Ala Gly
    475                 480                 485

CTC GCA CTT GTC TTT GCC GAG GCC CCC AGC CAG ATT CGC CAG GGT GTC      1599
Leu Ala Leu Val Phe Ala Glu Ala Pro Ser Gln Ile Arg Gln Gly Val
490                 495                 500                 505

CAG TCG GTC CAG CCC AAC AAT GCC TGG AAC CAG CTC TGC CCC AAG TAC      1647
Gln Ser Val Gln Pro Asn Asn Ala Trp Asn Gln Leu Cys Pro Lys Tyr
                510                 515                 520

GCG GCT CTT CCT CCC GAT TTG CAG T                                   1672
Ala Ala Leu Pro Pro Asp Leu Gln
            525
```

(2)SEQ ID NO:14 资料

(i)序列特征:

(A)长度:529 个氨基酸

(B)类型:氨基酸

(D)拓扑:线型

(ii)分子类型:蛋白质

(xi)序列描述:SEQ ID NO:14

```
Met Leu Ser Ser Ile Thr Leu Leu Pro Leu Leu Ala Ala Val Ser Thr
 1               5                  10                  15

Pro Ala Phe Ala Ala Val Arg Asn Tyr Lys Phe Asp Ile Lys Asn Val
            20                  25                  30

Asn Val Ala Pro Asp Gly Phe Gln Arg Ser Ile Val Ser Val Asn Gly
        35                  40                  45

Leu Val Pro Gly Thr Leu Ile Thr Ala Asn Lys Gly Asp Thr Leu Arg
    50                  55                  60

Ile Asn Val Thr Asn Gln Leu Thr Asp Pro Ser Met Arg Arg Ala Thr
65                  70                  75                  80

Thr Ile His Trp His Gly Leu Phe Gln Ala Thr Thr Ala Asp Glu Asp
                85                  90                  95

Gly Pro Ala Phe Val Thr Gln Cys Pro Ile Ala Gln Asn Leu Ser Tyr
            100                 105                 110

Thr Tyr Glu Ile Pro Leu Arg Gly Gln Thr Gly Thr Met Trp Tyr His
        115                 120                 125

Ala His Leu Ala Ser Gln Tyr Val Asp Gly Leu Arg Gly Pro Leu Val
    130                 135                 140

Ile Tyr Asp Pro Asn Asp Pro His Lys Ser Arg Tyr Asp Val Asp Asp
145                 150                 155                 160

Ala Ser Thr Val Val Met Leu Glu Asp Trp Tyr His Thr Pro Ala Pro
                165                 170                 175

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

Pro Val Pro Asp Ser Gly Leu Ile Asn Gly Lys Gly Arg Tyr Val Gly
        195                 200                 205

Gly Pro Ala Val Pro Arg Ser Val Ile Asn Val Lys Arg Gly Lys Arg
    210                 215                 220

Tyr Arg Leu Arg Val Ile Asn Ala Ser Ala Ile Gly Ser Phe Thr Phe
225                 230                 235                 240

Ser Ile Glu Gly His Ser Leu Thr Val Ile Glu Ala Asp Gly Ile Leu
                245                 250                 255

His Gln Pro Leu Ala Val Asp Ser Phe Gln Ile Tyr Ala Gly Gln Arg
            260                 265                 270

Tyr Ser Val Ile Val Glu Ala Asn Gln Thr Ala Ala Asn Tyr Trp Ile
        275                 280                 285

Arg Ala Pro Met Thr Val Ala Gly Ala Gly Thr Asn Ala Asn Leu Asp
    290                 295                 300

Pro Thr Asn Val Phe Ala Val Leu His Tyr Glu Gly Ala Pro Asn Ala
305                 310                 315                 320

Glu Pro Thr Thr Glu Gln Gly Ser Ala Ile Gly Thr Ala Leu Val Glu
                325                 330                 335
```

Glu Asn Leu His Ala Leu Ile Asn Pro Gly Ala Pro Gly Gly Ser Ala
340 345 350

Pro Ala Asp Val Ser Leu Asn Leu Ala Ile Gly Arg Ser Thr Val Asp
355 360 365

Gly Ile Leu Arg Phe Thr Phe Asn Asn Ile Lys Tyr Glu Ala Pro Ser
370 375 380

Leu Pro Thr Leu Leu Lys Ile Leu Ala Asn Asn Ala Ser Asn Asp Ala
385 390 395 400

Asp Phe Thr Pro Asn Glu His Thr Ile Val Leu Pro His Asn Lys Val
405 410 415

Ile Glu Leu Asn Ile Thr Gly Gly Ala Asp His Pro Ile His Leu His
420 425 430

Gly His Val Phe Asp Ile Val Lys Ser Leu Gly Gly Thr Pro Asn Tyr
435 440 445

Val Asn Pro Pro Arg Arg Asp Val Val Arg Val Gly Gly Thr Gly Val
450 455 460

Val Leu Arg Phe Lys Thr Asp Asn Pro Gly Pro Trp Phe Val His Cys
465 470 475 480

His Ile Asp Trp His Leu Glu Ala Gly Leu Ala Leu Val Phe Ala Glu
485 490 495

Ala Pro Ser Gln Ile Arg Gln Gly Val Gln Ser Val Gln Pro Asn Asn
500 505 510

Ala Trp Asn Gln Leu Cys Pro Lys Tyr Ala Ala Leu Pro Pro Asp Leu
515 520 525

Gln

```
   1 AGCGTCACACCAGACATCGGATGAAAACGGAAAGTGTATGCGCCATTTGACGTCTGCGGC    60

  61 AACCACTGTTCATCTCGCGAGCTAACATGGGCGACGTATAAGAAGAACGCGAGAATGGGC   120

 121 AGATTTCGATATCCCCTCTCGTCTCGGTTTTGGTCTCGGCTTGCCTCTAATGGCGCGCAC   180
                                                 M  A  R  T     4

 181 CACTTTCCTTGTCTCGGTTTCGCTCTTTGTTTCCGCTGTTCTTGCGCGCACCGTCGAGTA   240
   4   T  F  L  V  S  V  S  L  F  V  S  A  V  L  A  R  T  V  E  Y    24

 241 CGGCTTGAAGATTAGTGATGGGGAGATAGCTCCTGACGGTGTTAAGCGTAATGCGACTTT   300
  24   G  L  K  I  S  D  G  E  I  A  P  D  G  V  K  R  N  A  T  L    44

 301 GGgtacgcactccttgtaatccaacaattcaaggtttctgatgcttggtcagTAAATGGA   360
  44                                                 V  N  G     47

 361 GGGTATCCCGGTCCACTCATTTTTGCCAACAAGGGGGATACTCTCAAAGTCAAGGTCCAA   420
  47  G  Y  P  G  P  L  I  F  A  N  K  G  D  T  L  K  V  K  V  Q    67

 421 AACAAGCTCACGAATCCTGAGATGTATCGCACCACTTCCATCgtatgttcgttcgatatc   480
  67  N  K  L  T  N  P  E  M  Y  R  T  T  S  I                     81

 481 tactaatacatccgtcgctaaatatcttgtagCATTGGCACGGTCTCTTACAACATAGAA   540
  81                                 H  W  H  G  L  L  Q  H  R     90
```

图. 1A

```
541 ACGCCGACGACGACGGTCCTTCGTTCGTCACTCAGgtaggattctggaaggttggcctga   600
 90 N  A  D  D  D  G  P  S  F  V  T  Q                             102

601 actctctgttaaccgacaacccgatgtcaccagTGCCCGATTGTTCCACGCGAGTCGTAT   660
102                                   C  P  I  V  P  R  E  S  Y    111

661 ACTTACACCATACCTCTGGACGATCAAACCGGAACCTATTGGTACCATAGCCACTTGAGT   720
111  T  Y  T  I  P  L  D  D  Q  T  G  T  Y  W  Y  H  S  H  L  S   131

721 TCGCAATACGTTGATGGTCTTCGAGGCCCGCTGGTAATCTgtgagtatcttgacttgtct   780
131  S  Q  Y  V  D  G  L  R  G  P  L  V  I                       144

781 actgaaggcaacgagactaaaacaagcgtcgattcacagATGgttcgtctccccttttatt   840
144                                           Y                   145

841 tagctctggatcttcatttctcacgtaatacatgatagATCCCAAGGATCCTCACAGGCG   900
144                                       D  P  K  D  P  H  R  R   152

901 TTTGTATGATGTTGACGATGAGAAGACCGTCCTGATCATCGGTGACTGGTATCATGAATC   960
152   L  Y  D  V  D  D  E  K  T  V  L  I  I  G  D  W  Y  H  E  S   172

961 GTCCAAGGCAATCCTTGCTTCTGGTAACATTACCCGACAgtaagtgatacatgccggtcc  1020
172   S  K  A  I  L  A  S  G  N  I  T  R  Q                       185
```

图 1B

```
1021 cagaaaaattctctaaattcattttaattacagGCGACCGGTCTCTGCCACCATCAACGG 1080
 185                                    R  P  V  S  A  T  I  N  G  194

1081 CAAAGGTCGATTTGACCCTGACAACACTCCTGCCAACCCAGATACTCTGTACACCCTCAA 1140
 194   K  G  R  F  D  P  D  N  T  P  A  N  P  D  T  L  Y  T  L  K  214

1141 GGTCAAGCGAGGGAAGCGCTATCGTCTGCGTGTCATCAATAGCTCGGAGATCGCTTCGTT 1200
 214   V  K  R  G  K  R  Y  R  L  R  V  I  N  S  S  E  I  A  S  F  234

1201 CCGATTCAGTGTGGAAGGTCACAAGGTGACTGTGATTGCTGCCGATGGCGTCTCTACCAA 1260
 234   R  F  S  V  E  G  H  K  V  T  V  I  A  A  D  G  V  S  T  K  254

1261 ACCGTATCAGGTCGATGCGTTTGATATTCTAGCAGGACAGCGCATAGATTGCGTCgtaag 1320
 254   P  Y  Q  V  D  A  F  D  I  L  A  G  Q  R  I  D  C  V        272

1321 tgtcgtccgaacccacatctgagctcaagtgttgatacatgcgcgcttatagGTGGAGGC 1380
 272                                                   V  E  A  275

1381 GAACCAAGAACCCGACACATACTGGATCAACGCACCGCTGACCAACGTGCCCAACAAGAC 1440
 275   N  Q  E  P  D  T  Y  W  I  N  A  P  L  T  N  V  P  N  K  T  295

1441 CGCTCAGGCTCTCCTCGTTTATGAGGAGGATCGTCGGCCGTACCACCCTCCAAAGGGCCC 1500
 295   A  Q  A  L  L  V  Y  E  E  D  R  R  P  Y  H  P  P  K  G  P  315

1501 GTATCGCAAGTGGAGCGTCTCTGAGGCGATCATCAAGTACTGGAATCACAAGCACAAGCA 1560
 315   Y  R  K  W  S  V  S  E  A  I  I  K  Y  W  N  H  K  H  K  H  335
```

图. 1C

```
1561 CGGACGTGGTTTGCTGTCTGGACATGGAGGTCTCAAGGCTCGGATGATCGAGGGTAGCCA 1620
 335   G  R  G  L  L  S  G  H  G  G  L  K  A  R  M  I  E  G  S  H   340

1621 TCATCTGCATTCGCGCAGCGTCGTTAAGCGCCAGAATGAGACCACCACTGTTGTAATGGA 1680
 340   H  L  H  S  R  S  V  V  K  R  Q  N  E  T  T  T  V  V  M  D   350

1681 CGAGAGCAAGCTCGTTgtaagtaccatatttaaaagttggttgggtttcgaatacttatt 1740
 350   E  S  K  L  V

1741 tcaacttttcttagCCACTGGAATACCCCGGCGCTGCATGCGGGTCTAAACCTGCTGACC 1800
 350                  P  L  E  Y  P  G  A  A  C  G  S  K  P  A  D    365

1801 TCGTCTTGGATCTCACTTTTGGTTTGgtatgtagccaaatcgcccatatacaggatactg 1860
 365 L  V  L  D  L  T  F  G  L                                      374

1861 aatattgtttgtgcgtgtagAACTTTGCTACCGGGCACTGGATGATCAACGGTATCCCAT 1920
 374                        N  F  A  T  G  H  W  M  I  N  G  I  P    387

1921 ACGAGTCTCCCAAAATCCCCACATTGCTCAAGATCCTCACTGATGAGGACGGGGTTACCG 1980
 387 Y  E  S  P  K  I  P  T  L  L  K  I  L  T  D  E  D  G  V  T   407

1981 AGTCTGACTTgtatgttccctttcggtatcttcgtatgcgtgcactgactcgtgctggt 2040
 407 E  S  D  F                                                    411

2041 gggaatttagCACCAAGGAGGAGCACACAGTCATACTCCCGAAGAACAAATGCATCGAAT 2100
 411               T  K  E  E  H  T  V  I  L  P  K  N  K  C  I  E   427
```

图 1D

```
2101 TCAACATCAAGGGGAACTCGGGTATTCCCATTACGCACCCCGTACATCTTCACGGTgtaa 2160
 427 F  N  I  K  G  N  S  G  I  P  I  T  H  P  V  H  L  H  G       446

2161 gtgcatatcggatggtttacgatactaaggctcatcaacttttagCACACTTGGGATGT 2220
 446                                                H  T  W  D  V  451

2221 CGTACAATTTGGCAACAACCCACCCAATTATGTCAATCCTCCCCGTAGGGACGTGGTTGG 2280
 451   V  Q  F  G  N  N  P  P  N  Y  V  N  P  P  R  R  D  V  V  G  471

2281 CTCTACAGATGCGGGTGTGAGGATTCAGTTCAAGACCGACAATCCAGGACCGTGGTTCCT 2340
 471   S  T  D  A  G  V  R  I  Q  F  K  T  D  N  P  G  P  W  F  L  491

2341 GCACTGgtgcgtcggtccccatcgtccgttatggtttttctaatacgtcccattctattt 2400
 491   H  C                                                        493

2401 tagCCATATTGACTGGCATCTTGAGGAGGGTTTCGCAAgtgagtactgagacctaagtgc 2460
 493      H  I  D  W  H  L  E  E  G  F  A                          504

2461 tactcggctcattactgattaccgcatgtatgcgtctagTGGTGTTTGCTGAAGCGCCCG 2520
 504                                        M  V  F  A  E  A  P    511

2521 AAGCCGTCAAGGGAGGTCCAAAGAGCGTGGCCGTGGACTCTCAGTGGGAAGGGCTGTGTG 2580
 511 E  A  V  K  G  G  P  K  S  V  A  V  D  S  Q  W  E  G  L  C  531

2581 GCAAGTACGACAACTGGCTAAAATCAAATCCGGGCCAGCTGTAGGCGTATCGCAGCCACA 2640
 531 G  K  Y  D  N  W  L  K  S  N  P  G  Q  L  *                 545
```

图. 1E

2641 TTGGTGATGATTGAAAGTTGCATCTTGTTCCTATAACCGGCTCTTATATACGGGTGTCTC 2700

2701 CCAGTAAAGTCGTAGCCCAATTTCAGCCGAGACAGATATTTAGTGGACTCTTACTCTTGT 2760

2761 GTCCCATTGACGCACATCGTTGCATCAAACCTGCTTTTTATCGTCCCTCTTTGTAATTTG 2820

2821 TGTTGCTGTAATGTATCG 2838

图. 1F

```
   1 AAGCTTCGGCATGGATTGCATTTTGTATTGT                                    180

 181 AAACAAGTTACGAGAAAAACAATAGATCAGTTTTTGCCGAATCGGATGGCTTGAAACGGA      240

 241 AGTACCGATGGCCGATCCGAGTCGAATGAATTAACGCATCTGAAACGGGACCCTGAGTCG      300

 301 AGGCACCCGCCGGCCTTGGCCGTATAAGTCACTTGTCGCCAACTAGCACTTTTTCATTCC      360

 361 CCCTTTTCTTCTTCCTCGTCTTCTTCTTCTCTATGGCTCGGTCGACTACTTCACTCTTTG      420
   1                                   M  A  R  S  T  T  S  L  F       10

 421 CACTGTCTCTGGCCGCACCGGCCTTGGCTCGAGTCGTTGACTATGGGTTTGATGTGGCTA      480
  10 A  L  S  L  A  A  P  A  L  A  R  V  V  D  Y  G  F  D  V  A       30

 481 ATGGGGCAGTTGCTCCGGATGGTGTAACAAGGAACGCGGTTCTCGgtgagttagctgtaa      540
  30 N  G  A  V  A  P  D  G  V  T  R  N  A  V  L                     45

 541 gatggtgtatatgctggttgcctaacgggaatgtcagTCAATGGTCGCTTCCCTGGTCCA      600
  45                                          V  N  G  R  F  P  G  P   53

 601 TTGATCACCGCCAACAAGGGGGATACACTTAAAATCACCGTGCGGAATAAACTCTCCGAT      660
  53  L  I  T  A  N  K  G  D  T  L  K  I  T  V  R  N  K  L  S  D      73
```

图. 2A

```
661 CCAACTATGCGAAGGAGCACGACCATCgttagtacttcccctcatctgtcttgaaacttt   720
 73  P  T  M  R  R  S  T  T  I                                      82

721 ctcatctttttgaagCACTGGCACGGTCTGCTCCAACACAGGACGGCAGAAGAAGATGG   780
 82                  H  W  H  G  L  L  Q  H  R  T  A  E  E  D  G    97

781 CCCGGCCTTTGTAACCCAGgtatgccttatcctatcgctgctctgtccccgcgtccttcc   840
 97    P  A  F  V  T  Q                                           103

841 ctgactcgggcgattctagTGCCCGATTCCTCCGCAAGAATCGTACACCTATACGATGCC   900
103                     C  P  I  P  P  Q  E  S  Y  T  Y  T  M  P   117

901 GCTCGGCGAACAGACCGGCACGTATTGGTACCACAGCCACTTGAGCTCCCAGTATGTGGA   960
117   L  G  E  Q  T  G  T  Y  W  Y  H  S  H  L  S  S  Q  Y  V  D   137

961 CGGGTTGCGTGGGCCCATCGTTATTTgtaagtcttcatttaaccttattcttggctatgg  1020
137   G  L  R  G  P  I  V  I                                      145

021 ctgattgtgacgtcgtggttagATGgttcgtggcttccacaagaagtcagcagcccttga  1080
145                        Y                                      145

1081 agctaactttattccagACCCCCACGACCCGTACAGAAACTACTATGATGTCGACGACGA  1140
145                  D  P  H  D  P  Y  R  N  Y  Y  D  V  D  D  E   160

1141 GCGTACGGTCTTTACTTTAGCAGACTGGTACCACACGCCGTCGGAGGCTATCATTGCCAC  1200
160   R  T  V  F  T  L  A  D  W  Y  H  T  P  S  E  A  I  I  A  T   180
```

图. 2B

```
1201 CCACGATGTCTTGAAAACgtacgcgttaatccttctagctttctttccttgggtcacttt 1260
 180   H  D  V  L  K  T                                           185

1261 ctatcagGATCCCCGACTCGGGTACGATCAACGGCAAAGGCAAATACGATCCTGCTTCGG 1320
 185          I  P  D  S  G  T  I  N  G  K  G  K  Y  D  P  A  S    202

1321 CTAACACCAACAACACGACACTCGAGAACCTCTACACTCTCAAAGTCAAACGCGGCAAGC 1380
 202 A  N  T  N  N  T  T  L  E  N  L  Y  T  L  K  V  K  R  G  K    222

1381 GGTATCGCCTGAGGATTATCAACGCCTCGGCCATCGCTTCGTTCCGGTTCGGCGTGCAGG 1440
 222 R  Y  R  L  R  I  I  N  A  S  A  I  A  S  F  R  F  G  V  Q    242

1441 GCCACAAGTGCACGATCATCGAGGCTGATGGCGTCCTCACCAAACCGATCGAGGTCGATG 1500
 242 G  H  K  C  T  I  I  E  A  D  G  V  L  T  K  P  I  E  V  D    262

1501 CGTTTGATATTCTAGCAGGCCAGAGGTATAGCTGCATCgtaagtctacctatgccttgtt 1560
 262 A  F  D  I  L  A  G  Q  R  Y  S  C  I                       275

1561 gtggagataagaacctgactgaatgtatgcgctccaatagTTGAAGGCCGACCAAGATCC 1620
 275                                          L  K  A  D  Q  D  P    282

1621 TGATTCCTACTGGATAAATGCGCCAATCACAAACGTTCTCAACACCAACGTCCAGGCATT 1680
 282   D  S  Y  W  I  N  A  P  I  T  N  V  L  N  T  N  V  Q  A  L    302

1681 GCTAGTGTATGAAGATGACAAGCGTCCTACTCACTACCCCTGGAAGCCGTTTTTGACATG 1740
 302   L  V  Y  E  D  D  K  R  P  T  H  Y  P  W  K  P  F  L  T  W    322
```

图. 2C

```
1741 GAAGATATCAAATGAAATCATTCAGTACTGGCAGCACAAGCACGGGTCGCACGGTCACAA 1800
 322   K  I  S  N  E  I  I  Q  Y  W  Q  H  K  H  G  S  H  G  H  K   342

1801 GGGAAAGGGGCATCATCATAAAGTCCGGGCCATTGGAGGTGTATCCGGGTTGAGCTCCAG 1860
 342   G  K  G  H  H  H  K  V  R  A  I  G  G  V  S  G  L  S  S  R   362

1861 GGTTAAGAGCCGGGCGAGTGACCTATCGAAGAAGGCTGTCGAGTTGGCTGCTGCACTCGT 1920
 349   V  K  S  R  A  S  D  L  S  K  K  A  V  E  L  A  A  A  L  V   349

1921 TGCGGGTGAGGCCGAGTTGGACAAGAGGCAGAATGAGGATAATTCGACTATTGTATTGGA 1980
 349   A  G  E  A  E  L  D  K  R  Q  N  E  D  N  S  T  I  V  L  D   361

1981 TGAGACCAAGCTTATTgtaagtcccttaatttttttcggtgtcacggaagctaacccgcg 2040
 361   E  T  K  L  I                                                361

2041 taatagCCGTTGGTTCAACCTGGTGCACCGGGCGGCTCCAGACCAGCTGACGTCGTGGTC 2100
 361         P  L  V  Q  P  G  A  P  G  G  S  R  P  A  D  V  V  V   379

2101 CCTCTGGACTTTGGCCTCgtatgtggcttcttgttattcgtccggaatgcaaactgattt 2160
 379   P  L  D  F  G  L                                            385

2161 gggtgggctatagAACTTTGCCAACGGACTGTGGACGATAAACAATGTCTCCTACTCCCC 2220
 385                  N  F  A  N  G  L  W  T  I  N  N  V  S  Y  S  P   401

2221 TCCGGATGTCCCTACTCTCCTCAAGATCTTGACCGACAAAGACAAAGTCGACGCTTCTGA 2280
 401   P  D  V  P  T  L  L  K  I  L  T  D  K  D  K  V  D  A  S  D   421
```

图. 2D

```
2281 CTTgtaggttcctcttcttcttttcaaactagctactgacattaagtgaacgtcagCACG 2340
 421   F                                                      T   423

2341 GCCGATGAACACACGTATATTCTTCCAAAGAACCAAGTTGTCGAGTTGCACATCAAGGGA 2400
 423  A  D  E  H  T  Y  I  L  P  K  N  Q  V  V  E  L  H  I  K  G   453

2401 CAGGCTTTGGGAATCGTACACCCCCTTCATCTGCATGGCgtacgtctttctcacactgtt 2460
 453  Q  A  L  G  I  V  H  P  L  H  L  H  G                      466

2461 ccagctcctattctctaacacactcctgcgatagCATGCGTTCGACGTCGTCCAATTCGG 2520
 466                                    H  A  F  D  V  V  Q  F  G  475

2521 CGACAACGCTCCAAACTACGTGAACCCTCCGCGTAGGGATGTAGTAGGCGTAACTGATGC 2580
 475   D  N  A  P  N  Y  V  N  P  P  R  R  D  V  V  G  V  T  D  A  495

2581 TGGAGTCCGTATCCAGTTCAGAACCGATAACCCGGGCCCTTGGTTCCTCCATTGgtatgc 2640
 495   G  V  R  I  Q  F  R  T  D  N  P  G  P  W  F  L  H  C        513

2641 tcttcatctcccaccgcttgttctttacttatggtttaccttgcgatttagCCACATTGA 2700
 513                                                   H  I  D  516

2701 TTGGCACTTGGAAGAAGGATTTGCTAgtaagttattattcctattccgaagcatcgggga 2760
 516   W  H  L  E  E  G  F  A                                    524

2761 gatgctaaccaagggtgtgttttaagTGGTATTCGCCGAAGCGCCTGAAGATATCAAGAA 2820
 524                           M  V  F  A  E  A  P  E  D  I  K  K  536
```

图. 2E

```
2821 AGGCTCTCAGAGTGTCAAGCCTGACGGACAATGGAAGAAACTATGCGAGAAGTATGAGAA 2880
 536   G  S  Q  S  V  K  P  D  G  Q  W  K  K  L  C  E  K  Y  E  K  556

2881 GTTGCCTGAAGCACTGCAGTGAAGTTGCAGTTGTTTCCCATTCGGGAACTGGCTCACTAT 2940
 556   L  P  E  A  L  Q  *                                        562

2941 TCCTTTTGCATAATTCGGACTTTTATTTTGGGACATTATTGGACTATGCATTTGTTTGTC 3000

3001 ACACCGCGGAACTAAGCCGAATTC
```

图. 2F

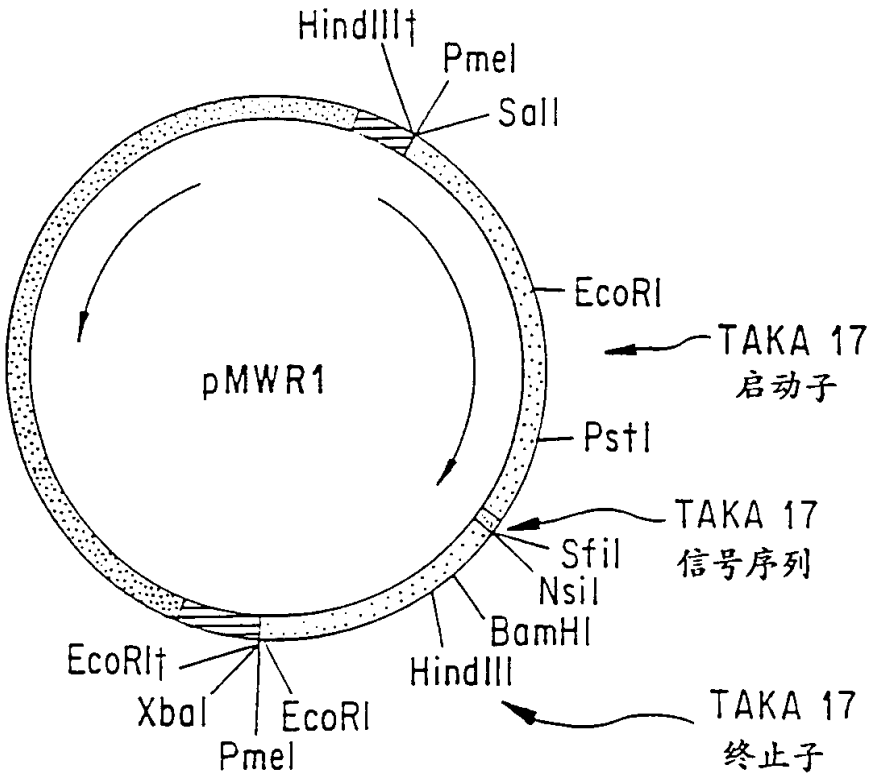
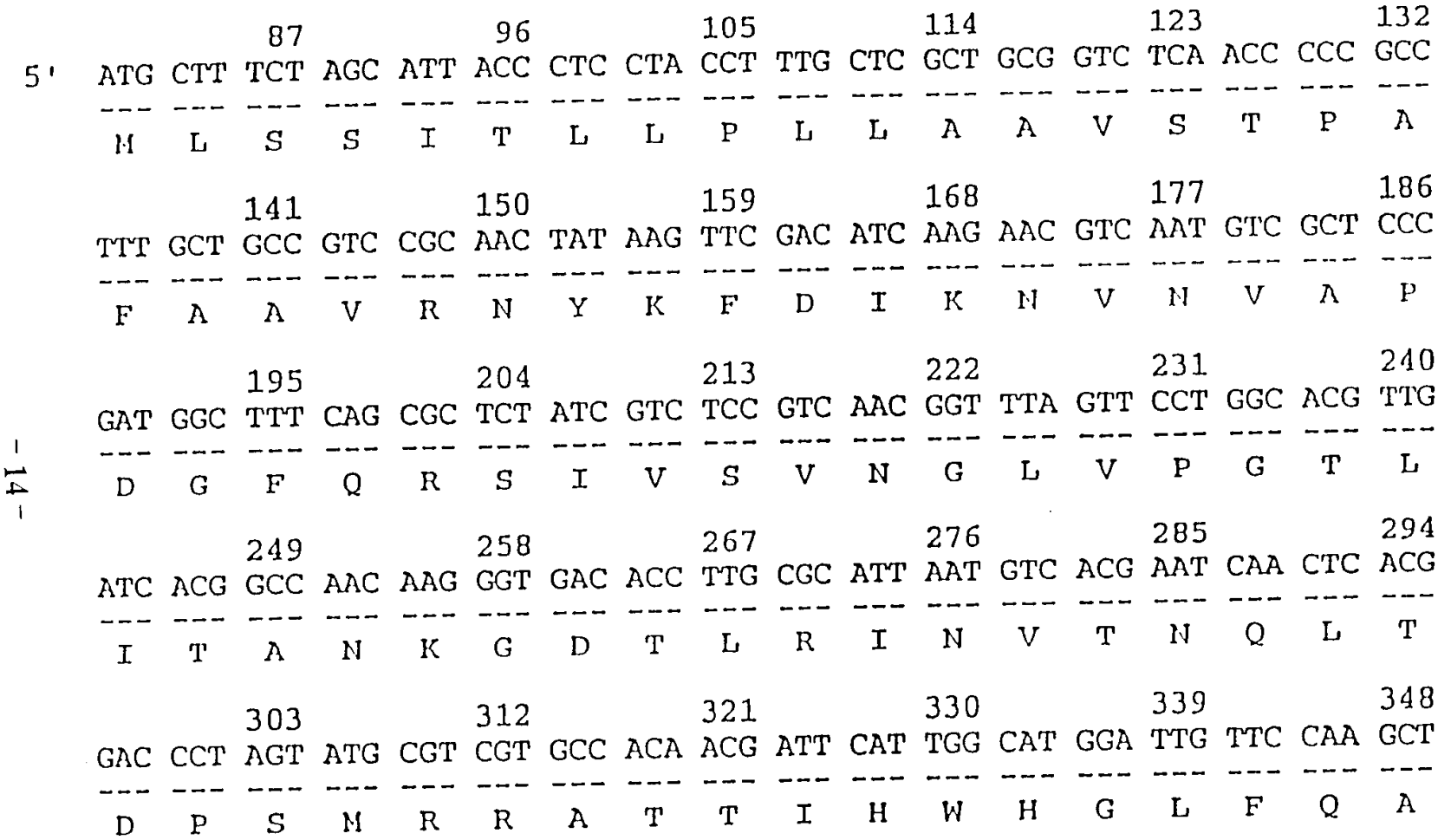

HindIII†
PmeI
SalI
EcoRI
TAKA 17
启动子
pMWR1
PstI
TAKA 17
信号序列
SfiI
NsiI
BamHI
HindIII
TAKA 17
终止子
EcoRI†
XbaI
EcoRI
PmeI

图. 3



```
                87          96          105         114         123         132
5'  ATG CTT TCT AGC ATT ACC CTC CTA CCT TTG CTC GCT GCG GTC TCA ACC CCC GCC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   L   S   S   I   T   L   L   P   L   L   A   A   V   S   T   P   A

                141         150         159         168         177         186
    TTT GCT GCC GTC CGC AAC TAT AAG TTC GAC ATC AAG AAC GTC AAT GTC GCT CCC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   A   A   V   R   N   Y   K   F   D   I   K   N   V   N   V   A   P

                195         204         213         222         231         240
    GAT GGC TTT CAG CGC TCT ATC GTC TCC GTC AAC GGT TTA GTT CCT GGC ACG TTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   G   F   Q   R   S   I   V   S   V   N   G   L   V   P   G   T   L

                249         258         267         276         285         294
    ATC ACG GCC AAC AAG GGT GAC ACC TTG CGC ATT AAT GTC ACG AAT CAA CTC ACG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   T   A   N   K   G   D   T   L   R   I   N   V   T   N   Q   L   T

                303         312         321         330         339         348
    GAC CCT AGT ATG CGT CGT GCC ACA ACG ATT CAT TGG CAT GGA TTG TTC CAA GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   P   S   M   R   R   A   T   T   I   H   W   H   G   L   F   Q   A
```

图. 4A

```
                357                 366                 375                 384                 393                 402
ACT ACC GCC GAC GAG GAT GGC CCC GCA TTC GTC ACG CAA TGC CCT ATT GCG CAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   T   A   D   E   D   G   P   A   F   V   T   Q   C   P   I   A   Q

                411                 420                 429                 438                 447                 456
AAT TTG TCC TAT ACA TAC GAG ATC CCA TTG CGC GGC CAA ACA GGA ACC ATG TGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   L   S   Y   T   Y   E   I   P   L   R   G   Q   T   G   T   M   W

                465                 474                 483                 492                 501                 510
TAT CAC GCC CAT CTT GCG AGT CAA TAT GTC GAT GGA TTG CGA GGC CCT TTG GTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Y   H   A   H   L   A   S   Q   Y   V   D   G   L   R   G   P   L   V

                519                 528                 537                 546                 555                 564
ATC TAT GAT CCA AAC GAC CCA CAC AAG TCG CGC TAC GAC GTG GAT GAT GCG AGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   Y   D   P   N   D   P   H   K   S   R   Y   D   V   D   D   A   S

                573                 582                 591                 600                 609                 618
ACA GTA GTC ATG CTT GAG GAC TGG TAC CAT ACT CCG GCA CCC GTT CTA GAA AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   V   V   M   L   E   D   W   Y   H   T   P   A   P   V   L   E   K
```

图. 4B

```
                627                 636                 645                 654                 663                 672
CAA ATG TTC TCG ACT AAT AAC ACC GCT CTG CTC TCT CCT GTT CCG GAC TCG GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   M   F   S   T   N   N   T   A   L   L   S   P   V   P   D   S   G

                681                 690                 699                 708                 717                 726
CTT ATC AAT GGC AAA GGG CGC TAT GTG GGC GGT CCC GCA GTT CCC CGG TCA GTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   I   N   G   K   G   R   Y   V   G   G   P   A   V   P   R   S   V

                735                 744                 753                 762                 771                 780
ATC AAC GTA AAA CGT GGG AAA CGA TAT CGC TTG CGC GTA ATC AAC GCT TCT GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   N   V   K   R   G   K   R   Y   R   L   R   V   I   N   A   S   A

                789                 798                 807                 816                 825                 834
ATC GGG TCG TTT ACC TTT TCG ATC GAA GGA CAT AGT CTG ACT GTC ATT GAG GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   G   S   F   T   F   S   I   E   G   H   S   L   T   V   I   E   A

                843                 852                 861                 870                 879                 888
GAT GGG ATC CTG CAC CAG CCC TTG GCT GTT GAC AGC TTC CAG ATT TAC GCT GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 D   G   I   L   H   Q   P   L   A   V   D   S   F   Q   I   Y   A   G
```

图. 4C

```
          897         906         915         924         933         942
CAA CGC TAC TCT GTC ATC GTT GAA GCC AAC CAA ACC GCC GCC AAC TAC TGG ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   R   Y   S   V   I   V   E   A   N   Q   T   A   A   N   Y   W   I

          951         960         969         978         987         996
CGT GCA CCA ATG ACC GTT GCA GGA GCC GGA ACC AAT GCA AAC TTG GAC CCC ACC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   A   P   M   T   V   A   G   A   G   T   N   A   N   L   D   P   T

         1005        1014        1023        1032        1041        1050
AAT GTC TTT GCC GTA TTG CAC TAC GAG GGA GCG CCC AAC GCC GAA CCC ACG ACG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   V   F   A   V   L   H   Y   E   G   A   P   N   A   E   P   T   T

         1059        1068        1077        1086        1095        1104
GAA CAA GGC AGT GCT ATC GGT ACT GCA CTC GTT GAA GAG AAC CTC CAT GCG CTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   Q   G   S   A   I   G   T   A   L   V   E   E   N   L   H   A   L

         1113        1122        1131        1140        1149        1158
ATC AAC CCT GGC GCT CCG GGC GGC TCC GCT CCC GCA GAC GTT TCC CTC AAT CTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   N   P   G   A   P   G   G   S   A   P   A   D   V   S   L   N   L
```

图. 4D

```
        1167        1176        1185        1194        1203        1212
GCA ATT GGG CGC AGC ACA GTT GAT GGG ATT CTT AGG TTC ACA TTT AAT AAC ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   I   G   R   S   T   V   D   G   I   L   R   F   T   F   N   N   I

        1221        1230        1239        1248        1257        1266
AAG TAC GAG GCT CCT TCG TTG CCC ACG CTC TTG AAG ATT TTG GCA AAC AAT GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   Y   E   A   P   S   L   P   T   L   L   K   I   L   A   N   N   A

        1275        1284        1293        1302        1311        1320
AGC AAT GAC GCC GAT TTC ACG CCA AAT GAG CAC ACT ATC GTA TTG CCA CAC AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   N   D   A   D   F   T   P   N   E   H   T   I   V   L   P   H   N

        1329        1338        1347        1356        1365        1374
AAA GTT ATC GAG CTC AAT ATC ACC GGA GGT GCA GAC CAC CCT ATC CAT CTC CAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   V   I   E   L   N   I   T   G   G   A   D   H   P   I   H   L   H

        1383        1392        1401        1410        1419        1428
GGC CAT GTG TTT GAT ATC GTC AAA TCA CTC GGT GGT ACC CCG AAC TAT GTC AAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   H   V   F   D   I   V   K   S   L   G   G   T   P   N   Y   V   N
```

图. 4E

```
        1437        1446        1455        1464        1473        1482
CCG CCA CGC AGG GAC GTA GTT CGT GTC GGA GGC ACC GGT GTG GTA CTC CGA TTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 P   P   R   R   D   V   V   R   V   G   G   T   G   V   V   L   R   F

        1491        1500        1509        1518        1527        1536
AAG ACC GAT AAC CCA GGC CCA TGG TTT GTT CAC TGC CAC ATT GAC TGG CAC TTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   T   D   N   P   G   P   W   F   V   H   C   H   I   D   W   H   L

        1545        1554        1563        1572        1581        1590
GAG GCT GGG CTC GCA CTT GTC TTT GCC GAG GCC CCC AGC CAG ATT CGC CAG GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   A   G   L   A   L   V   F   A   E   A   P   S   Q   I   R   Q   G

        1599        1608        1617        1626        1635        1644
GTC CAG TCG GTC CAG CCC AAC AAT GCC TGG AAC CAG CTC TGC CCC AAG TAC GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   Q   S   V   Q   P   N   N   A   W   N   Q   L   C   P   K   Y   A

        1653        1662
GCT CTT CCT CCC GAT TTG CAG T 3'
--- --- --- --- --- --- --- -
 A   L   P   P   D   L   Q
```

图. 4F

01-04-26

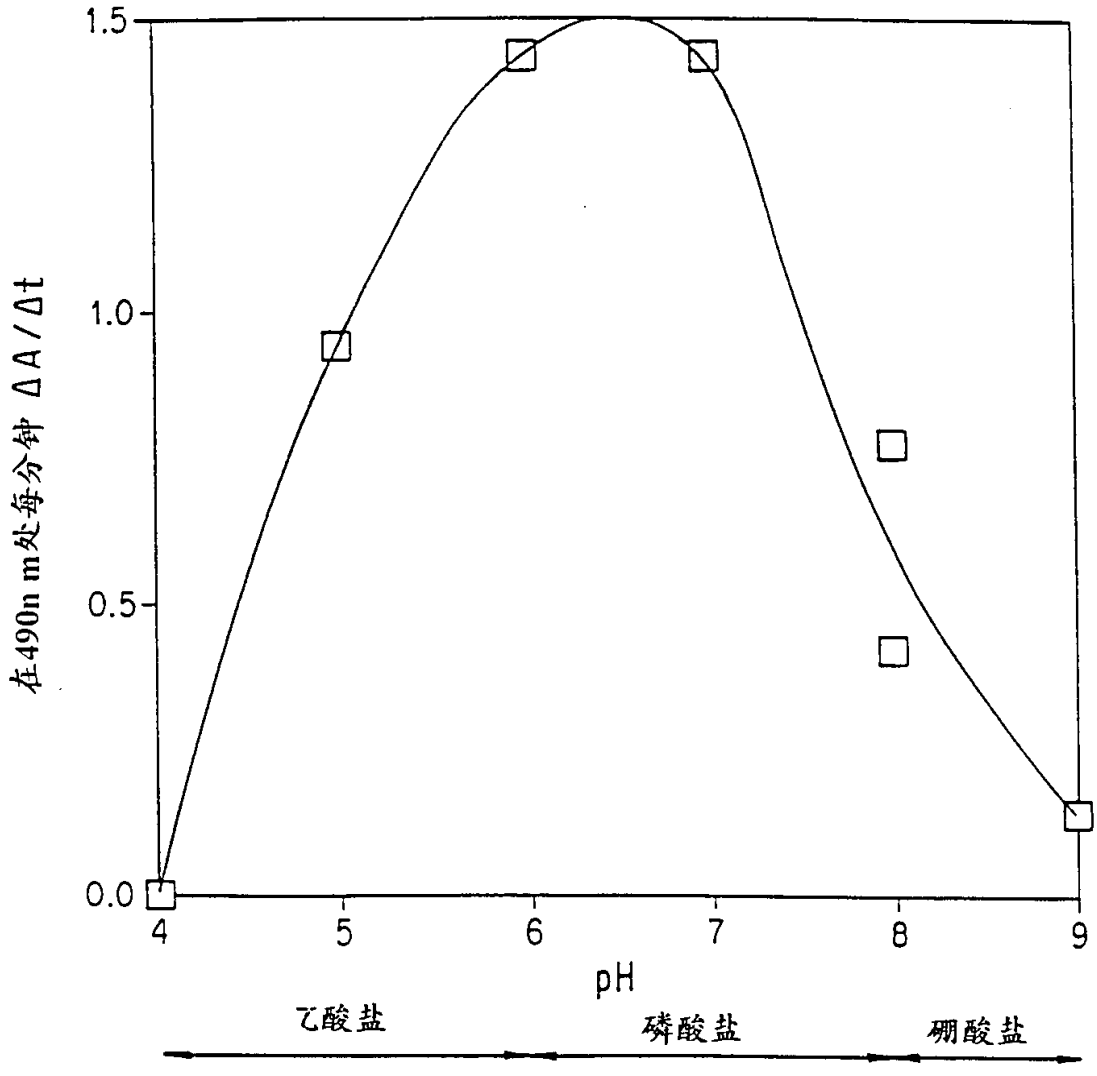


图. 5

01.04.26

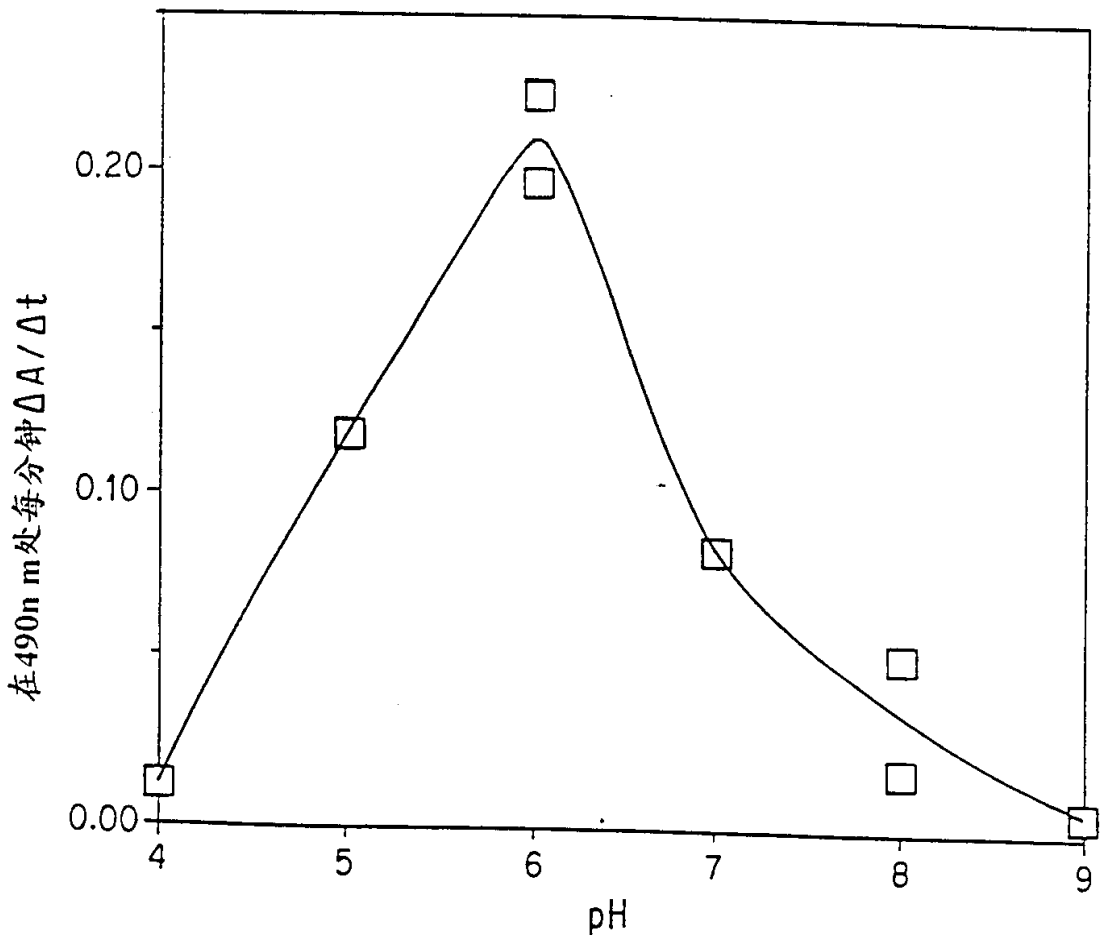


图. 6