(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105106929 A
(43)申请公布日 2015.12.02

(21)申请号 201510600024.5
(22)申请日 2006.09.26
(30)优先权数据
60/720,540 2005.09.26 US
(62)分案原申请数据
200680043903.1 2006.09.26
(71)申请人 绿色细胞有限公司
地址 美国马里兰
(72)发明人 W·J·劳瑟 W-S·约 K·洛里克 M·佩奇斯 N·J·劳特
(74)专利代理机构 北京北翔知识产权代理有限公司 11285
代理人 张广育 姜建成
(51)Int.Cl.
A61K 38/02(2006.01)
A61K 8/64(2006.01)
C12N 15/82(2006.01)
A01H 5/00(2006.01)

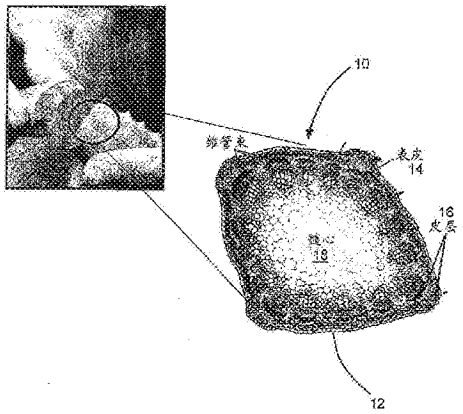
权利要求书1页 说明书21页
序列表30页 附图12页

(54)发明名称

用于生产蛋白的转基因芦荟植物及其相关的方法

(57)摘要

本发明可以提供转基因的芦荟植物和用于芦荟植物转化的重组构建体，本发明的方面可以被应用于其他单子叶植物。所述重组构建体可以包含一个或多个编码哺乳动物蛋白的DNA序列，和至少一个能够在芦荟植物中指导所述重组蛋白表达的启动子。本发明还提供用于构建和产生一个转基因芦荟植物的方法。本发明包括同时用数个目的基因转染一个芦荟植物的方法。本发明的所述芦荟植物的制备方法可以提供经济地并更安全地大量生产某些生物活性化合物的可能并且更容易为不富裕国家所接受，所述化合物包括用于疾病治疗、诊断和预防的生物药剂。所述芦荟植物产生方法还可以制备用作化妆品的蛋白。



CN 105106929 A

1.一种治疗组合物,所述组合物包含:

一种有生物活性的外源蛋白,所述蛋白由在一个芦荟植物中引入的编码序列表达,并且至少部分的所述蛋白是从所述芦荟植物的芦荟细胞被易位到芦荟叶的凝胶中;和

一种或多种天然的和治疗性芦荟蛋白或碳水化合物。

2.一种产生生物活性蛋白的方法,所述方法包含:

制备一种DNA构建体,所述DNA构建体含有一个启动子、一个目的编码序列、一个终止序列和一个易位序列;

将所述DNA构建体克隆到一个芦荟植物的基因组中;

增殖所述含所述DNA构建体的芦荟植物;并且

提取一种通过所述DNA构建体表达的目的蛋白,其中至少部分的所述蛋白是从所述芦荟植物的芦荟细胞被易位到芦荟叶的凝胶中。

3.根据权利要求2的方法,所述方法进一步包含诱导编码目的蛋白的基因转录的步骤。

用于生产蛋白的转基因芦荟植物及其相关的方法

[0001] 本申请为分案申请，其原申请的申请日为2006年9月26日，申请号为200680043903.1(国际申请号PCT/US2006/037756)，名称为“用于生产蛋白的转基因芦荟植物及其相关的方法”。

[0002] 相关专利申请的交叉引用

[0003] 本专利申请要求2005年9月26日提交的名称为“TRANSGENIC ALOE PLANTS FOR PROCUTION OF PROTEINS AND RELATED METHODS”并且指定序列号为60/720,540的美国临时申请的优先权和权益，该申请全文通过引用的方式纳入本文。

技术领域

[0004] 本发明涉及转基因单子叶植物，并且更具体地涉及转基因芦荟植物，以及用于产生转基因芦荟植物的方法和组合物，和从所述转基因芦荟植物提取蛋白的方法。

背景技术

[0005] 具有生物活性的蛋白有持续增长的市场，许多的所述生物活性蛋白被应用于治疗目的。目前已有超过160种市售的基于蛋白的药物。未来的两年，另外还有大约50种有希望被批准。现在对治疗性蛋白生产的需求已经超过了工业的产能。为了克服这个问题，预计所述工业的产能需要提高4-5倍。然而，这些治疗性蛋白的生产设备昂贵，并且其建造需要较长的时间。因此，需要更加低成本的生产方法，并且需要所述方法能够缩短达到大规模生产所需要的时间。

[0006] 可以使用一些基于动物的蛋白生产方法。但是，这些方法经常会引入疾病造成的健康风险。这样的风险可能来自于疾病的交叉感染，所述疾病可以同时感染所述动物和例如人类患者的最终使用者。因此，存在一个对生产方法的需求，所述方法需要消除所述生产生物和所述终端使用者之间的交叉感染的可能性。

[0007] 另外，许多现有的生产方法需要大量的处理，以从所述动物或产生治疗性蛋白的其他宿主生物中提取所述治疗性蛋白，并且使所述化合物处于患者可以利用的状态。经过纯化以后，可以将所述蛋白与佐剂或其他载体材料结合以稳定所述蛋白并且使患者可以利用。然而，涉及对治疗性蛋白处理的提取、纯化、在其他溶剂中的再悬浮的过程是复杂和繁琐的，并且可能不利于在对治疗有普遍需求的发展中国家中使用。因此，需要一个简化生产方法，所述方法可以消除或减少治疗性蛋白的提取后处理。

发明内容

[0008] 根据本发明的组合物和方法可以解决上述的许多的需求和缺陷，并且将会提供另外的改进和优点，本领域技术人员通过阅读本发明公开的内容将会理解这些改进和优点。

[0009] 在一方面，本发明可以提供一种稳定地整合目标基因的转基因芦荟植物。在另一方面，可以提供新的载体和构建体，以将目标DNA序列整合进芦荟植物的基因组中。所述目标序列可以编码一种生物活性的蛋白。所述生物活性的蛋白可以是干扰素、免疫球蛋白、淋

巴因子、生长因子、激素、血液因子、组织相容性抗原、酶、化妆品用蛋白和其他哺乳动物蛋白，或其他目标蛋白。在某些方面，所述目标蛋白是人类蛋白。根据本发明的芦荟植物可以将目标基因转录并翻译成目标蛋白。在一方面，至少一些所述的目标蛋白可以易位到所述芦荟叶的中央部分。在其他方面，所述的目标蛋白可以包含一个信号序列以便于其易位到所述芦荟叶的中央部分。在其他方面，可以提供从芦荟植物上分离单细胞的新方法。仍然在其他的方面，本发明可以提供新的方法，所述方法用于将载体整合到芦荟植物中，并且再生这样的转基因芦荟植物。

[0010] 根据本发明的产生哺乳动物蛋白的转基因芦荟植物可以提供用于生产目标蛋白的、更加经济的可行的选择，所述的目标蛋白例如各种生物活性的蛋白和化妆品用蛋白。在一方面，所述的目标蛋白可以定位于和 / 或集中在所述芦荟叶的凝胶中。所述的目标蛋白的这种定位可以简化从所述植物中移出所述蛋白。在这方面，所述的目标蛋白可以与位于所述芦荟叶中央部分的非变性凝胶被共提取。因此，本发明可以提供一种转基因植物，与例如烟草和玉米的大部分的转基因植物相比，通常可以更容易地从所述转基因植物获得目标蛋白。而且，本发明可以提供有效的蛋白分离方法。仍然在其他的方面，所述的目标蛋白可能不是特别地位于所述芦荟植物中。然而，所述芦荟植物的解剖学和生理学仍然可以提供生产目标蛋白的某些其他的优点，本领域的技术人员通过阅读本公开内容将可以理解这些优点。

[0011] 与在细菌和酵母中的产生目标蛋白的常规方法相比，芦荟植物有各种优点。芦荟植物 10 的所述优点可以包括以某些方式加工蛋白的能力，所述方式不适合于简单的单细胞细菌和酵母，并且是产生具有所需形式的蛋白必须的。例如，这种加工可以包括化学修饰（例如通过糖基化）和某些蛋白的折叠。而且，与基于动物细胞的其他蛋白生产方法相比，芦荟植物生产可以提供显著的成本优势、规模优势和减少可能对人有害的污染的危险。

[0012] 来自根据本发明的转基因芦荟植物的芦荟叶的中心部分的蛋白修饰凝胶从所述芦荟植物取出就可以被直接应用。这可以避免相对复杂的和昂贵的从天然植物材料中提取目标蛋白的过程。在某些方面，从所述叶中提取的凝胶不需要蛋白的提取或处理就可以被直接应用。所述凝胶可以以髓心的形式，所述髓心是从转基因芦荟植物的断裂叶的脱落部分机械地提取的。

[0013] 本发明可以为具有生物活性和 / 或具有化妆品用途的人和其他哺乳动物蛋白的生产提供经济、可行的选择。

[0014] 通过阅读本公开内容，本领域技术人员将会理解本发明的其他改进和优点。

附图说明

[0015] 图 1 显示芦荟叶的横断面解剖的一个实例；

[0016] 图 2 显示产生胼胝体组织的示例性方法；

[0017] 图 3 用图表概述了产生转基因芦荟植物的各个步骤；

[0018] 图 4 用图表显示包含序列信息的质粒载体的组件；

[0019] 图 5 列出来自玉米的泛素启动子的序列信息，并且突出显示了所述泛素启动子区域；

[0020] 图 6A 和 6B 显示根据本发明的方面的质粒载体系统的构建；

[0021] 图 7A 至 7C 显示根据本发明的方面的另一个质粒载体系统的构建；

[0022] 图 8 显示根据本发明的方面的另外一个质粒载体系统的构建；和

[0023] 图 9 显示根据本发明的方面的整合系统的构建。

[0024] 所有的附图的显示仅仅是为了易于解释本发明的基本教导；在阅读如下的说明之后，所述附图关于各种实施方案的数量、位置、序列、关系和组分的扩展将能够被本领域的技术人员解释或者将在所述技术人员的理解范围内。而且，在阅读如下的说明之后，用于实践本发明的各种方案、工具和组合物将在本领域的技术人员的理解范围内。

具体实施方式

[0025] 本说明书中使用的，“一”或“一种”可能指一种或多种。如在权利要求中使用的，当与词语“包含”一块使用的时候，词语“一”或“一种”可能指一种或多于一种。如本文使用的“另一种”可能指至少第二种或更多。

[0026] 本文使用的，一种“转化的芦荟细胞”指一种用稳定整合的、非天然的、重组 DNA 转化（例如农杆菌介导的转化或者通过使用包被有重组 DNA 的微粒或其他方法轰击）的植物细胞。本发明的一种转化的芦荟细胞可以是一种原始转化的以一种微生物或者作为一种子代植物细胞形式存在的植物细胞，所述子代植物细胞再生成分化的组织，例如再生成具有稳定整合的、非天然的重组 DNA 的转基因芦荟植物 10，或者衍生自子代转基因芦荟植物 10 的种子或花粉。

[0027] 本文使用的“转基因芦荟植物”指一种基因组被稳定整合的重组 DNA 改变的芦荟植物。一个转基因的芦荟植物 10 包括从一个原始转化的芦荟细胞再生的芦荟植物，以及来自转化的芦荟植物 10 的后代或杂交的子代转基因芦荟植物。

[0028] 本文使用的“重组 DNA”指经过基因工程改造并且在细胞外构建的 DNA，包括含有天然存在的 DNA、cDNA、合成的 DNA 和 / 或其他 DNA。

[0029] 本文使用的“启动子”指起始转录的调控 DNA。“植物启动子”是能够起始芦荟细胞中的转录的启动子，不管它是否来自于芦荟细胞，例如，众所周知农杆菌启动子在芦荟细胞中起作用。因此，植物启动子包括从植物、植物病毒和细菌（例如农杆菌和短根瘤菌）获得的启动子 DNA。受发育控制的启动子的实例包括优先在在特定组织中起始转录的启动子，所述组织例如叶、根或种子。这些启动子指的是“组织优选的”。仅仅在特定组织中起始转录的启动子指的是“组织特异性的”。一种“细胞类型”特异性的启动子主要地驱动在一种或多种器官中的特定细胞类型内的表达，例如在根或叶中的维管细胞。“可诱导的”或“可抑制的”启动子是一种受环境控制的启动子。可能影响诱导型启动子转录的环境条件的实例包括缺氧条件、或者特定的化学试剂、或者光的存在。组织特异性的、组织优选的、细胞类型特异性的和诱导型的启动子构成“非组成型”启动子的类。“组成型的”启动子是一种在大部分条件下有活性的启动子。所述启动子可能包含增强子或其他元件，所述其他元件影响转录的起始、转录的起始位点、转录的水平、转录的末端位点或者所产生的核糖核酸的任何的后处理。

[0030] 本文使用的术语“遗传构建体”被限定为一段包含人工排列的至少两个 DNA 片段的 DNA 序列，所述 DNA 序列是为了产生一种转基因的芦荟植物。在一个具体的实施方案中，一个片段是一个调控序列，并且另一个片段编码一种基因产物。

[0031] 本文使用的“有效连接(operably linked)”的意思是在一个DNA构建体中联合两段或多段DNA片段,以至于一段的功能(如编码蛋白的DNA)被另一段调控(例如启动子、终止序列等)。

[0032] 本文使用的术语“转录”被限定为从一个DNA模板生成一个RNA分子。

[0033] 本文使用的术语“翻译”被限定为从一个RNA模板生成一个多肽。

[0034] 本文使用的“表达的”指产生的,例如,当一个蛋白的对应的DNA被转录成mRNA时,所述mRNA被翻译成一个蛋白,所述蛋白在植物细胞中表达。

[0035] 本发明可以提供表达各种目标蛋白的新转基因芦荟植物10、可以提供产生转基因芦荟植物10的方法和组合物、可以提供从所述转基因芦荟植物10中提取蛋白的方法,并且可以提供来自所述转基因芦荟植物10的组分与目标蛋白的新组合物。根据本发明产生的目标蛋白可以包含一个分泌信号,以帮助所述蛋白在所述髓心18中或者在芦荟叶12的髓心中蓄积。在一方面,这种蓄积可以至少部分地出现在转基因芦荟植物10的叶的musalegenous细胞中。

[0036] 所述髓心18或芦荟叶12的浆包含所述叶的组分,当从所述芦荟叶12提取所述组分的时候,所述叶的组分通常至少部分地指的是“凝胶”。本文使用的“凝胶”将指所提取的髓心18和其他相关物质,在从所述芦荟叶12中提取髓心18的时候所述相关物质伴随于所述髓心18,而不管后续处理的程度。根据本发明的一个或多个方面,来自所述芦荟叶12的凝胶可以有修饰组分。在一个具体方面,所述凝胶的组分可以被修饰为包含至少一种外源性的蛋白组分,并被称为“蛋白修饰凝胶”。提取含有相关的目的蛋白的蛋白修饰凝胶可以提供容易获得的蛋白和/或有效的蛋白分离的方法。根据本发明产生和/或提取的蛋白修饰凝胶不需要进行蛋白提取就可以直接被应用。

[0037] 出于示例性目的,本发明的方面通常显示在图1至8中。本发明可以提供转基因植物、组合物,以及通常可以被应用于百合家族的芦荟属的方法。典型地,对本发明的描述参考了在物种中的应用,所述物种选自于芦荟(Aloe vera)(库拉索芦荟)、开普芦荟(Aloe ferox)、木立芦荟(Aloe arborescence)。所述芦荟属通常包括一组大的无茎、丛生、多质单子叶植物。这些植物通常指用于本公开内容目的的芦荟植物10。

[0038] 芦荟植物产生的各种化合物已经被应用于药用目的。当存在于一种蛋白修饰凝胶中的时候,这些化合物可以补充根据本发明的转基因的芦荟植物10的生物活性蛋白的药学性质。图1显示转基因的芦荟植物10的芦荟叶12的横断面。所述转基因的芦荟植物10的叶包含一种容易提取的凝胶状蛋白混合物、碳水化合物和所述凝胶中包含的水,所述凝胶主要来自于髓心18。该凝胶状混合物主要位于芦荟叶12的中心区域。已经显示所述凝胶中的各种化合物具有多种药用性质和应用。在一方面,所述凝胶和/或髓心18可以稳定转基因的芦荟植物10产生的、并且位于所述凝胶和/或髓心18中的转基因蛋白。

[0039] 图1特别地显示表皮14、皮层或叶肉16和髓心或髓18。所述表皮14形成所述叶的外层细胞。在本发明的一个方面,一个转基因的芦荟植物10可以包含位于一种或多种这些组织中的芦荟细胞,所述细胞暂时地或稳定地整合一种遗传的构建体,所述遗传构建体被所述芦荟细胞表达。所述皮层16包含富含叶绿体,以及维管束、木质部和韧皮部的细胞。所述髓心18是海绵状的薄壁组织并且至少部分地代表所述凝胶,所述薄壁组织几乎仅由大的皮层细胞组成,所述凝胶可以有助于整合或蓄积一种或多种目标转基因蛋白。

[0040] 所述表皮 14 通常由单外层细胞组成。就在所述表皮 14 的下面是包含维管束网络的皮层 16。所述维管束的外支撑通常由鞘细胞提供。所述维管束由三种常见的管状结构组成：木质部、韧皮部和相关的大的中柱鞘管。所述木质部将水和矿物质从所述植物的根运送到叶。所述韧皮部在整个植物中运输淀粉和其他合成的物质。所述中柱鞘管包含一种胶乳或树液，所述胶乳或树液含有非常多的轻泻剂蒽醌，特别是芦荟素。所述蒽醌吸收太阳的紫外线，并且防止所述芦荟叶 12 的中央部分过热，所述芦荟叶 12 通常的功能是作为所述芦荟植物的水贮藏器官。所述维管束的中柱鞘部分附着到所述表皮 14 上，而所述维管束的剩余部分突出到所述髓心 18 中。所述芦荟叶 12 的最里面和主要部分是髓心 18，所述髓心 18 至少部分地组成所述凝胶。为了提取凝胶，通常认为所述表皮 14 和皮层 16 可以包括含有凝胶的鞘。所述提取的凝胶（包含大量的髓心 18）通常是稠的并粘的物质，所述物质在历史上为了医学的目的已经被局部应用于皮肤上，例如作为一种烧伤和创伤的治疗物。

[0041] 所述凝胶通常的功能是作为所述芦荟植物的物质的蓄积器。所述皮层 16 通常合成多种碳水化合物和糖蛋白，所述碳水化合物和糖蛋白是所述芦荟植物需要的。过量合成的碳水化合物通常被转运到髓心 18，与水和某些矿物质一块储存。所述碳水化合物被韧皮部管转运到髓心 18 的皮层 16 的细胞的大液泡。然后，因为渗透压的原因水被所述碳水化合物吸收，从而使所述髓心 18 作为所述芦荟植物的水贮藏器官。

[0042] 通过沿其纵轴破碎芦荟叶 12 并且挤压所述叶，将周围的鞘中的所述凝胶通过所述断裂面挤出来，可以相对容易地提取所述凝胶。可以将其他的物质与所述凝胶一块从所述叶组织中提取出来。

[0043] 方法概述

[0044] 根据一个或多个本发明的转基因的芦荟植物 10 通常包含一种或多种 DNA 构建体，所述 DNA 构建体稳定地整合在所述植物中以表达一种或多种目标蛋白。根据一个或多个本发明产生一个转基因的芦荟植物可以涉及各种新的组合物和方法。典型地，开发一种或多种 DNA 构建体以在所述转基因的芦荟植物 10 中表达一种或多种蛋白。在一方面，单独一种构建体可以表达一种 mRNA。在另一方面，单独一种构建体可以表达多种 mRNA。在又另一方面，多种构建体可以产生多种 mRNA。每种 mRNA 可以产生一种或多种多肽。

[0045] 为了将所述构建体引入到所述芦荟植物中，芦荟细胞或未分化的胼胝体组织通常被应用，如图 2 的一般性地显示。所述芦荟细胞和胼胝体组织典型地来自于根或叶分生组织，或者来自于一个芦荟植物的种子。使用许多如图 3 中图表显示的技术和 / 或构建体，将所述 DNA 构建体引入所述芦荟细胞中。通过阅读本公开内容，本领域的技术人员将会知晓某些技术。可以使用各种技术筛选已整合有所需构建体的芦荟细胞或胼胝体组织。典型地，所述 DNA 构建体将包含一种选择性标记。一旦将目标构建体稳定地引入所述芦荟细胞或胼胝体组织，通常从所述芦荟细胞或胼胝体组织产生芦荟的小植物。可以发育成为包含所述 DNA 构建的活的转基因的芦荟植物 10 的芦荟小植物可以组成型地表达或者诱导性地表达目标的一种或多种目标蛋白。依赖于所产生的特定的蛋白和 / 或是否存在所述蛋白相关的信号序列，目标蛋白可以在所述转基因的芦荟植物 10 的局部，或者在所述转基因的芦荟植物 10 的全株中普遍存在。依赖于所述特定的构建体和 / 或相关的目标蛋白，然后可以通过无性或者有性的方式繁殖所述转基因的芦荟植物 10。可以使用许多技术以从所述转基因的芦荟植物 10 中分离所述蛋白。然后可以进一步地分离和 / 或进一步地处理所述蛋白。

这些处理可以包括酶修饰、化学修饰、掺入合适的佐剂，这些处理以及其他处理对本领域的技术人员来说通过阅读本公开内容后是可以知晓的。在一个方面，可以处理所述蛋白，将它从一种非活性（前体）的形式转变成一种活性形式，以进行所述特定蛋白的具体的应用，。

[0046] 细胞的分离

[0047] 如图 2 中的示例性序列所示，在整合目标构建体之前，通常从一个芦荟植物中分离芦荟细胞或者芦荟细胞群。通常基于其再生能力来选择用于 DNA 构建体的芦荟细胞类型。可以使用来自芦荟植物的茎分生组织或根分生组织的分生细胞或者来自芦荟种子的胚的芦荟细胞。然而，所述芦荟植物的许多部分保持再生或形成胼胝体组织的能力，并且也可以使用所述部分。通常可以使用许多技术分离所述细胞，本领域的技术人员通过阅读本公开内容将会知晓所述技术。典型地，使用解剖刀机械地从所述芦荟植物分离所述细胞。或者，其他机械的或非机械的技术可以被应用于分离目标细胞，这是本领域的技术人员通过阅读本公开内容后可以知晓的。一旦分离了合适的芦荟细胞或者芦荟细胞群，通常使所述芦荟细胞在培养基中生长以形成胼胝体组织。虽然某些技术不需要所述芦荟细胞首先生长成胼胝体组织，但是所述胼胝体组织提供了未分化芦荟细胞的集合的来源，所述未分化芦荟细胞保持产生转基因的芦荟植物 10 的能力。所述胼胝体组织通常在固体培养基上生长。然而，所述胼胝体组织也可以被置于液体培养基中，并且悬浮生长。

[0048] 所述分生细胞可以从芦荟植物的根尖或叶中分离。分生芦荟细胞也可以从顶端分生组织中分离。典型地，用于分离所述组织的芦荟植物是年轻健康的芦荟植物。通常选择的所述芦荟植物不高于八 (8) 英寸，并具有六 (6) 个或少于 6 个的出生茎。通常通过切割成段在一个芦荟植物的茎上形成创伤。所述创面可以刺激所述芦荟细胞生长。通常，所述胼胝体将主要在所述切表面上开始生长。为了从芦荟叶 12 分离芦荟细胞，通常从新生枝的远端切段。然后将所述段部分置于生长培养基上。枝顶端分生组织源自于距新生芦荟植物的底端一英寸处的切割。将所述芦荟叶 12 从所述切面去除，并且防止所述段。在所述“切口”中发现所述分生组织芦荟细胞。通过它们在培养皿上不断生长—形成胼胝体组织或再生枝的能力“选择”或“分离”它们。一旦分离所述分生组织芦荟细胞，就将它们培养在合适的培养基上，并且处于促进胼胝体组织形成的条件下。

[0049] 从种子分离所述胚的芦荟细胞通常包括去除所述种皮以暴露所述胚，并且从所述胚机械地剥离目标芦荟细胞。所述芦荟种子经典型地灭菌处理，并且除去所述种子的外壳。为了从芦荟植物或者芦荟胚机械地移出所述芦荟细胞，通常可以使用解剖刀。一旦从所述种皮中分离胚的芦荟细胞，就将它们培养在合适的培养基上，并且处于促进胼胝体组织形成的条件下。

[0050] 当需要胼胝体组织时，所述分离的芦荟细胞通常在有利于胼胝体组织的条件下生长，这是本领域的技术人员通过阅读本公开内容可以知晓的。典型地，将所述分离的芦荟细胞在合适的固体营养培养基中放置并生长，以在转化之前形成直径大约为 1cm 大小的胼胝体组织。芦荟细胞通常生长在摄氏 23-26 度下。合适的培养基包括基础固体或液体培养基，所述培养基将通常包含补充的无机营养素—常量元素（例如氮、硫、磷、钙、镁和钾）和微量元素（例如铁、硼、钴、铜、碘、锰、钼和锌）；包括糖（蔗糖或麦芽糖）和维生素以及辅因子（硫胺素、烟酸、生物素、吡哆素、尤其是肌醇）的有机营养素；氨基酸（例如脯氨素和酪蛋白水解产物）；以及主要是生长素的源（通常是 (NAA), 1- 萘乙酸；(IAA), 吲哚 -3- 醋

酸；或 (2,4-D)，2,4- 二氯苯氧代乙酸）和细胞分裂素的源（通常为 (BAP)6- 苄氨基嘌呤）的生长调节剂。这些培养基和维生素通常是以预配制的组合物的形式市售，所述组合物具有各种浓度的无机营养素和维生素，并且已知有 MS 培养基 (Murashige-Skoog)、Gamborg 培养基或 Chu N6 培养基。另外，在培养皿 (Petri dishe) 上生长的细胞通常需要固体基质载体，例如在所述生长培养基中加入琼脂。

[0051] DNA 构建体的形成

[0052] 合适的 DNA 构建体通常被引入到由分离的芦荟细胞衍生的胼胝体组织，以通过形成转基因的芦荟植物 10 产生目标蛋白。也可以将 DNA 构建体引入分离的芦荟细胞或成熟的芦荟细胞，这是本领域的技术人员通过阅读本公开内容后可以知晓的。为了增殖所述构建体并且将其引入所述芦荟细胞，DNA 构建体通常被整合进入载体（例如质粒或病毒）中。一种示例性质粒载体可以包括由 Invitrogen(Carlsbad,California) 生产的商品名为 pZEr0 的质粒，出于示例性目的显示在图 4 的图表中，并且，例如可以包含该质粒的一种或多种功能组件。

[0053] 根据本发明的 DNA 构建体可以包括能够在芦荟细胞中起作用的启动子序列、编码目的蛋白的序列、终止序列和能够在芦荟细胞中起作用的翻译起始和终止位点。当正确地配置并组合这些组件的时候，它们可以在芦荟细胞中启动 DNA 的转录和 mRNA 的翻译，以及它们各自的终止。所述 DNA 构建体还可以包括分泌信号。另外，一些目标蛋白，例如干扰素，可以包含一个天然的或合成的结构域，所述结构域指引干扰素的分泌。当所述蛋白离开所述细胞的时候这些分泌信号通常会被切下来，或者在某些情况下这些分泌信号可以保持在所述蛋白上而基本上不影响所述蛋白的功能。可以将分泌信号加到各种需要分泌信号来进行易位的蛋白上。这些分泌信号可以来自于各种植物的分泌序列。所述 DNA 构建体或一种相关载体也可以包含至少一种可选标记。在一方面，所述 DNA 构建体可以包含用于所述载体在细菌中的增殖的第一可选标记，和一种用于在芦荟细胞中生长的可选标记。另外，所述 DNA 构建体可以包含其他的调控元件，所述元件也是为了在所述芦荟植物 10 中的特定芦荟细胞中表达。其他构建体元件可以包含额外的调控元件，例如增强转录的 5’端引导序列和内含子、3’端非翻译区（例如多腺苷酸化信号和位点）、运输肽的 DNA。目标 DNA 序列（例如编码目标治疗蛋白的序列）的扩增可以通过多聚酶链反应来实现（参见编号为 4,683,202 和 5,928,906 的美国专利，均通过引用纳入本文）。通过阅读本公开内容，本领域的技术人员将会明白，使用本领域的技术人员已知的方法可以修饰所述 DNA 构建体以改变它们的功能，所述 DNA 构建体包括启动子序列、调控序列、稳定序列、靶向序列和 / 或终止序列。

[0054] 如上文所述，通常可以使用一种在芦荟细胞中起作用的启动子。所述启动子通常包含调控蛋白和分子可以结合的遗传元件。这些蛋白通常包括 RNA 聚合酶和其他转录因子。所述启动子可以在一个功能位置和 / 或方向上与核酸序列可操作地连接，以控制该序列的转录起始和 / 或表达。可以将所述启动子与或者不与一个“增强子”连接，所述增强子是指一种顺式作用调控序列，所述调控序列涉及一段核酸序列的转录激活。

[0055] 自然地，使用一种启动子和 / 或增强子是重要的，所述启动子和 / 或增强子在表达必需的芦荟植物 10 的至少一种细胞类型中可以有效地指导所述 DNA 片段表达。那些分子生物学领域的技术人员通常知晓启动子、增强子和细胞型组合在蛋白表达中的用途，例如参见 Sambrook et al.(1989)，通过引用的方式纳入本文中。所述使用的启动子可以是组成

型的、组织特异性的、诱导型的和/或在合适的条件下起作用的启动子，以指导所述引入的DNA片段的高水平的表达，这对于大规模生产重组蛋白和/或肽是有利的。所述启动子可以是异源的或者内源的。所述DNA构建体通常需要能够在芦荟细胞中起作用的转录和翻译起始和终止调控信号。可以使用大量的调控转录起始的序列。控制转录起始的DNA序列可以来自于农杆菌、病毒或植物。来自花椰菜花叶病毒的35S病毒转录起始区域(35S-CaMV)可以被用于芦荟植物10。可以被应用于芦荟植物10的植物启动子还可以包括来自各种单子叶或双子叶植物的核酮糖-1,5-二磷酸羧化酶(RUBISCO)小亚基启动子或者来自玉米的泛素启动子。可以使用其他合适启动子，并且这是本领域的技术人员通过阅读本公开内容后可以知晓的。如果需要诱导型调控，那么结构域可以来自不同的来源，使得来自一个来源的调控区域与来自另一个来源的RNA多聚酶结合结构域组合。表达的调控可以是在转基因的芦荟植物的10发育的特定阶段、在所述转基因的芦荟植物10的特定部分(如根、叶、种子、花、树液)中，或者可以是在植物部分和发育阶段的组合中。对一个特定发育阶段或组织的表达调控可能需要其他的DNA元件，这是本领域的技术人员通过阅读本公开内容后是可以知晓的。

[0056] 许多在植物细胞中有活性的启动子已经在文献中有所描述，并且可能被应用于芦荟细胞中。它们包括存在于植物基因组中的启动子，以及其他来源的启动子，包括：根瘤土壤杆菌的根瘤诱导质粒上携带的胭脂碱合酶(NOS)启动子和章鱼碱合酶(OCS)启动子；和例如花椰菜花叶病毒的花椰菜花叶病毒启动子。另外，在各种参考文献中也已经鉴定了各种其他的启动子，所述参考文献包括但不限于编号为5,858,742和5,322,938的美国专利，所述专利公开了来自于花椰菜花叶病毒(CaMV35S)的组成型启动子版本；编号为5,641,876的美国专利，所述专利公开了一种水稻肌动蛋白启动子；公布号为2002/0192813A1的美国专利申请，所述申请公开了在有效植物表达载体设计中有用的5',3'和内含子元件；序列号为09/757,089的美国专利申请，所述申请公开了一种玉米叶绿体醛缩酶启动子；序列号为08/706,946的美国专利申请，所述申请公开了一种水稻谷蛋白启动子；序列号为09/757,089的美国专利申请，所述申请公开了一种玉米醛缩酶(FDA)启动子；和序列号为No.60/310,370的美国专利申请，所述申请公开了一种烟胺合酶启动子，所有这些专利和专利申请通过引用的方式纳入本文。这些和大量的在植物细胞中发挥功能的其他启动子可以被应用于转基因芦荟植物中，用于表达目标治疗蛋白。

[0057] 作为一个具体的实例，可以使用来自玉米的所述泛素启动子(SEQ.ID.NO.44)。所述来自玉米的泛素启动子是一个大的元件，接近2kb长，它由至少三段普通区域组成。所述的启动子的序列在图5中给出，并且标出了特别相关的特征。所述泛素启动子的第一段位于最5'端，并包含基质结合区(MAR)，所述MAR是与组蛋白和其他核蛋白相互作用的元件，并且作为"伸向环外的"侧翼序列，使之更易于接近所述细胞转录装置。它们也协助将转录元件彼此分开，这对于防止在一个位置起始的转录"通读"到第二个序列是重要的。这可以有利于减少形成反义信使的危险。

[0058] 所述第二段包含增强子元件和真正的启动子。所述增强子元件结合转录因子，所述转录因子负责指导所述转录装置(pol II复合体)结合启动子并且起始化转录。虽然"启动子"具体地指与所述核心转录起始装置相互作用必需的必要DNA元件，但是所述术语启动子更广泛地被应用于包括涉及转录起始的所有的DNA元件，并且在这种情况下，所述"泛素

启动子”被泛指目标克隆基因 5’端的 2kb 区域或其修饰。

[0059] 第三段包含一段大约 1kb 的内含子。内含子是转录的基因的区域，所述区域被通过剪接的步骤从最终翻译的信使 (mRNA) 中去除。也已经发现，通过备选的增强子元件以及通过促进 RNA 信使从核到细胞质的易位（部分地与剪接相关的过程）增加蛋白的翻译，内含子可以直接地影响基因表达。然而，在转基因系统中存在内含子不一定总是导致 RNA 表达或者蛋白翻译水平的提高。在某些方面，所述泛素内含子可以被修饰以减少所述载体的总大小，并且以评估它们对所述芦荟中的转基因表达的作用。

[0060] 在本发明的其他方面，可能需要在所述芦荟植物的绿色组织中优先表达。这种用途需要的启动子可以包括那些来自于如拟南芥核酮糖 -1,5- 二磷酸羧化酶 (Rubisco) 小亚基 (Fischhoff et al.(1992)Plant MoI Biol.20:81-93)、醛缩酶和丙酮酸正磷酸双激酶 (PPDK)(Taniguchi et al.(2000)Plant Cell Physiol.41(1):42-48) 的基因的启动子。

[0061] 如前所述，可以改变所述启动子以包含多个“增强子序列”，以协助增强基因表达。这样的增强子是本领域中已知的。通过在这样的构建体中包含一个增强子序列，可以增强所选蛋白的表达。虽然这些增强子通常存在于一个在真核细胞中有功能的启动子的转录起始的 5’端，但是也可以将它们插入到所述编码序列的上游 (5’) 或下游 (3’)。在某些情况下，这些 5’端增强元件是内含子。作为增强子特别有用的是水稻肌动蛋白 1（参见编号为 5,641,876 的美国专利）和水稻肌动蛋白 2 基因的 5’端内含子、玉米乙醇脱氢酶基因内含子、玉米热休克蛋白 70 基因内含子（编号为 5,593,874 的美国专利）和玉米皱缩 1 基因。

[0062] 根据本发明的方面的构建体可以包含一个 3’端元件，所述元件通常包含一个多腺苷信号和位点。众所周知的 3’端元件包括那些来自于根癌农杆菌基因的 3’端元件，如 nos 3’、tml 3’、tmr 3’、tms 3’、ocs 3’、tr73’，例如在编号为 6,090,627 的美国专利中公开的，通过引用的方式纳入本文；来自如小麦 (Triticum aesevitum) 热休克蛋白 17(Hspl7 3’)、小麦泛素基因、小麦果糖 -1,6- 二磷酸酶基因、水稻谷蛋白基因、水稻乳酸脱氢酶基因和水稻 β 微管蛋白基因的植物基因的 3’端元件，所有这些元件都在美国公布的专利申请 2002/0192813A1 中公开，该文献通过引用的方式纳入本文；和豌豆 (Pisum sativum) 核酮糖二磷酸羧化酶基因 (rbs 3’)，以及来自宿主植物内部的基因的 3’端元件。

[0063] 可以被整合进所述 DNA 构建体中的结构基因编码一种目标蛋白。所述结构基因可以是一个哺乳动物基因或哺乳动物基因的一部分。目标结构基因可以编码干扰素、免疫球蛋白、淋巴因子、生长因子、激素、血液因子、组织相容性抗原、酶或其他蛋白。出于示例性目的，干扰素 α -2 的序列被作为 SEQ. ID. NO. 33 列出。结构基因也可以编码标记蛋白，例如来自水母的绿色萦光蛋白 (GFP)。可以修饰所述结构基因的 DNA 序列以使其在芦荟植物 10 中高水平表达。所述芦荟植物 10 的密码子偏好可能不同于用于分离所述结构基因的原始物种的密码子偏好。可以改造天然的结构基因以优化所述芦荟植物 10 中编码的蛋白的产生。此外，可以改造所述结构基因的 DNA 序列以在芦荟植物 10 中提供合适的糖基化，所述糖基化不影响所述蛋白的结构。

[0064] 可以通过能够在芦荟植物 10 中起作用的各种转录和翻译终止序列提供转录和翻译的终止。在一方面，所述终止序列可以包括来自于农杆菌的胭脂碱合酶 (NOS) 基因的序列。在另一方面，所述终止序列可以衍生自芦荟本身的基因和蛋白的终止序列。

[0065] 在所述 DNA 构建体中经常整合一个标记基因。利用所述标记基因，可以从所有不

包含所述标记基因的细胞群中选择包含目标基因结构的细胞。标记基因可以包括酶或其他蛋白，所述其他蛋白提供卡那霉素、氯霉素、G418 和庆大霉素等抗性。如果以例如农杆菌作为一种载体，那么可能还需要其他特异性的 DNA 序列。如果所述 DNA 被定位于特定的细胞类型，那么可以有其他区域。

[0066] 如上文所述，所述 DNA 构建体也可以包括一种分泌信号。构建体还可以包含编码一种信号肽的易位序列，所述信号肽可以定位所述治疗蛋白，用于将所述蛋白从产生它的细胞中移出。在文献中已经鉴定了多种信号序列，所述信号序列在植物中发挥作用，更具体地说是在如芦荟植物的单子叶植物中发挥作用。这些信号序列可以被可操作地整合进入所述构建体中或者目标基因中。在一个方面，可以使用来自于水稻 (Oryza sativa) 的 α 淀粉酶分泌序列 (SEQ.ID NO.29)。对这个信号序列的特征的描述参见 "The alpha-amylase genes in Oryza sativa"MoI Gen Genet.,1990April ;221(2):235-44，它的全文通过引用的方法纳入本文。或者，一些目标基因——例如干扰素——可以包含一个结构域（天然的或合成的），所述结构域指导干扰素的分泌。所述分泌信号在所述蛋白离开细胞时被切掉，或者在某些情况下，可以被所述蛋白保留而基本上不影响所述蛋白的功能。分泌信号可以被加到各种需要分泌信号进行易位的蛋白上。这些分泌序列可以衍生自各种植物分泌序列。

[0067] 图 6A 至 6B 显示构建体的示例性集合，所述构建体用于从芦荟植物产生蛋白。用 pzUI(TGC9) 标记所述质粒构建体。TGC9 包含启动人干扰素 α-2 和 pzUEK(TGC12) (SEQ.ID.NO.38) 表达的全长的泛素启动子、启动包含 eGFP（增强的绿色荧光蛋白）(SEQ.ID.NO.31) 和卡那霉素抗性蛋白的融合蛋白表达的全长的泛素启动子。所述融合蛋白将两个蛋白（选择和显示）的筛选性质合并为一。

[0068] 如图所示，使用每条都包含两个不同限制性内切酶位点（引物序列见表 I) 的 PCR 引物，通过 PCR 扩增人干扰素 α-2 基因产生所述质粒 pzI(TGC8)。所述扩增的干扰素 (IFN) 基因具有 5'端的 PstI 和 BamHI 侧位点，以及 3'端的 SacI 和 XhoI 侧位点。所述 PCR 生成的 IFN 然后作为一个 PstI/XhoI 片段克隆进 pZErO 载体以产生所述构建体 pzI(TGC8)，并且为后续的克隆引入所述 BamHI 和 SacI。这个构建体包含全长的 IFNα 基因和其信号序列，负责指导 IFN 蛋白从所述细胞的分泌。这显示在滑动图 (slide)3 上。对这些 PCR 生成的克隆测序以确定不存在任何的 DNA 突变。

[0069] 如图所示，通过将 IFNα-2 基因克隆到所述全长的泛素启动子的下游产生所述质粒载体 pzUI(TGC9)。将所述 IFN 的 PstI/XhoI 片段从 pzI(TGC8) 中释放，并且与 pzU(TGC1) 连接以产生所述构建体 pzUI(TGC9)。这在滑动图 3 上显示。

[0070] 如图所示，使用包含 BamHI(5'端）和 XhoI(3'端）限制性内切酶位点（引物序列见表 I) 的引物进行 eGFP 基因的 PCR 扩增。这个 PCR 产生的 eGFP 基因没有终止密码子（以最终使一种 eGFP-kan 融合蛋白可以被表达）。因此，翻译不在所述 eGFP 序列末端终止。用限制性内切酶 BamHI 和 XhoI 消化所述 PCR 扩增的片段，并且连接到 pZErO 的 BamHI 和 XhoI 位点中，产生 pzEnostop(TGC10)。这显示在滑动图 3 上。对这些 PCR 生成的克隆测序以确定不存在任何的 DNA 突变。

[0071] 如图所示，pzEK(TGC11) 形成一种 eGFP 基因和卡那霉素抗性基因之间的融合构建体。将 eGFP 以 BamHI/XhoI 片段形式从 pzEnostop(TGC10) 释放，并且与 pzK(TGC4) 连接以产生所述构建体 pzEK(TGC11)。所述 eGFP 基因被克隆到阅读框内，并且在 kan/nos(SEQ.

ID.NO.35) 的 5’端。

[0072] 如图所示，包含 eGFP/kan/nos 盒的来自 pzEK(TGC11) 的 BamHI/XbaI 片段被克隆到 pzUI(TGC9) 的 BamHI 和 XbaI 位点，以使 eGFP/kan 融合基因可以表达。所述 IFN 序列因此被去除。这个新构建体 (pzUEK(TGC 12)) 从全长的泛素启动子（在滑动图 4 中描述）开始表达所述 eGFP/kan 融合蛋白。该融合蛋白保持每个单独的蛋白的特性，并且因此明显同时具有显示标记和选择性试剂双功能。

[0073] 图 7A 至 7C 显示一个其他示例性构建体集合，所述构建体用于从芦荟植物产生目标蛋白。所示载体使两种基因可以作为一个转录子表达，虽然所述转录子仍然被翻译成两个不同的蛋白。为了实现这个目的，通过插入 IRES 元件（内部核糖体进入序列）(SEQ.ID.NO.34) 将两个基因彼此分开。所述 IRES 元件可为核糖体附着和翻译提供一个第二内部位点。所述第二位点使转录子克隆到所述 IRES 元件的 3’端，以独立地翻译。这些载体，即 pzUSEIrK(TGC 18)(SEQ.ID.NO.43) 和 pzUIIrK(TGC19)(SEQ.ID.NO.39) 被分步骤地构建，最终使 eGFP 或 IFN（分别地）与所述卡那霉素抗性标记同时表达。

[0074] 如图所示，通过首先使用包含 BamHI(5’端）和 SacI(3’端）限制性内切酶位点（引物序列见表 I) 的引物进行 eGFP 基因的 PCR 扩增来产生 pzEnostart(TGC13) 载体。该 PCR 产物作为一个 BamHI/SacI 片段被克隆进 pZEr0。在这个构建体中的所述 eGFP 基因缺少 ATG 翻译起始位点，因为该蛋白是作为与所述 α 淀粉酶信号序列（见构建体 pzSE,TGC14) 的融合体表达的。这显示在滑动图 5 上。对这些 PCR 生成的克隆测序以确定不存在任何 DNA 的突变。

[0075] 如图所示，合成所述 α 淀粉酶基因的信号序列的一条链，然后通过使用 PCR（引物序列见表 I) 进行填补生成双链。然后，用限制性内切酶 PstI 和 BamHI 消化并克隆进 pzEnostart(TGC13)。这显示在滑动图 5 上。所述信号序列被克隆到阅读框中，eGFP 的 5’端，并且为信号序列 eGFP 融合体提供翻译起始位点。对产生的 pzSE(TGC14) 克隆测序以确定不存在任何的 DNA 突变。

[0076] 如图所示，构建所述载体 pzUSE(TGC15) 以将 α 淀粉酶信号序列 (ss)-eGFP 融合体与具有泛素启动子相连接。来自 pzSE(TGC14) 的包含所述 ss-eGFP 的基因盒被作为一个 PstI/SacI 片段克隆到 pzUI(TGC9) 中。这替代所述 IFN 基因盒（通过用限制性内切酶 PstI 和 sacI 消化从 pzUI(TGC9) 释放）产生 pzUSE(TGC15)。这显示在滑动图 5 上。

[0077] 如图所示，使用包含 SacI(5’端）和 XhoI(3’端）限制性内切酶位点的引物进行 PCR 扩增从 pIRES2-EGFP(Clontech,BD Biosciences) 产生 IRES 元件。所扩增的 IRES 元件作为一个 SacI/XhoI 片段被克隆到 pZEr0 中以形成 pzIr(TGC16)。这显示在滑动图 6 上。对这些 PCR 生成的克隆测序以确定不存在任何的 DNA 突变。

[0078] 如图所示，然后将来自 pzIr(TGC16) 的 IRES 元件作为一个 SacI/XhoI 片段克隆到 pzK(TGC4) 中以形成 pzIrK(TGC17)。因此，定位所述 kan/nos 基因盒的 IRES 上游。

[0079] 如图所示，所述 IRES-kan/nos 基因盒作为 SacI/XbaI 片段被克隆到 pzUSE(TGC15) 中以形成 pzUSEIrK(TGC18)。这显示在滑动图 6 上。该构建体将 eGFP 作为与所述 α 淀粉酶信号序列的融合体表达，定位所述 eGFP 以从所述细胞中分泌。这使得可视地监视蛋白在所述转化的植物内的运输和聚积成为可能。该载体还表达所述卡那霉素抗性蛋白（由所述 IRES 内部元件翻译），并且使得可以在转基因植物中选择。

[0080] 如图所示，所述 IRES-kan/nos 基因盒可以作为 SacI/XbaI 片段被克隆到 pzUI(TGC9) 中以形成 pzUIIrK(TGC19)。这显示在滑动图 7 上。该构建体表达 IFNα-2，它具有用于分泌的信号序列。该载体还表达所述卡那霉素抗性蛋白（由所述 IRES 内部元件翻译），并且可以在转基因植物中选择。TGC18 和 TGC19 都以单个的转录子形式表达两个基因，省去了对第二启动子的需求，并且因此缩小了每个载体的总大小。这是重要的，因为缩小转染的构建体的总大小会增加这些元件转移到核中的效率，导致了稳定的整合以及转基因植物的选择。单启动子系统还会降低破坏侧翼基因的表达的风险或者其他可能最终影响转基因表达的情况。

[0081] 上面列出的质粒可以用其他目标编码序列替代所述 α 干扰素。如本领域的技术人员知晓的，这些替代可以在所述质粒中的相同位点或者其他位点。

[0082] 根据本发明的方面，也可以将一种凝血酶原编码序列 (SEQ. ID. NO. 36) 整合进一个质粒载体，如图 8 中所图示显示的。凝血酶原是凝血酶的前体蛋白。它被活化的因子 X(Factor X) 在 2 位切割以释放活化的凝血酶，所述凝血酶是一种凝结蛋白，可以将可溶性的血纤蛋白原转化成不溶的纤维蛋白原带。使用 PCR 引物（引物序列见表 I）扩增所述凝血酶原基因盒 (1874 碱基对）。利用限制性内切酶 BamHI 将所述 eGFP 基因盒从 pzUSEIrK(TGC18) 载体上释放。随后将所产生的 5'突出粘性末端用绿豆核酸酶除去以形成平端。使用小牛肠磷酸酶 (CIP) 通过在摄氏 50 度下孵育 2 个小时进行所述平端的去磷酸化，之后在摄氏 15 度下与所述凝血酶原基因盒连接过夜。这产生 pzUSTIrK(TGC20)(SEQ. ID. NO. 42)。对 PCR 生成的克隆测序以确定不存在任何的 DNA 突变。

[0083] 根据本发明的方面，也可以将一种皮离蛋白 (Dermcidin)(DCD) 编码序列 (SEQ. ID. NO. 30) 整合进一个质粒载体，如图 8 中所图示显示的。皮离蛋白是一种最近报道的宽光谱抗菌肽，被发现在汗腺中组成型的表达。该蛋白分泌到汗液中，并且被转运到表皮的表面。所述 PCR 扩增的 DCD 基因盒具有侧翼的 BamHI 限制性内切酶的位点（引物序列见表 I）。所述 BamHI 消化的 DCD 基因盒与原来被 eGFP 占据的 pzUSEIrK(TGC18) 的 BamHI 位点连接，形成 pzUSDIrK(TGC21) 载体 (SEQ. ID. NO. 40)。对 PCR 生成的克隆测序以确定不存在任何的 DNA 突变。

[0084] 根据本发明的方面，也可以将一种人生长激素 (hGH) 编码序列 (SEQ. ID. NO. 32) 整合进一个质粒载体，也如图 8 中所图示显示的。所述 hGH 基因盒通过 PCR 扩增产生，具有侧翼的 BamHI 限制性内切酶的位点（引物序列见表 I）。然后将所述 BamHI 消化的 hGH 基因盒与原来被 eGFP 占据的 pzUSEIrK(TGC18) 的 BamHI 位点连接，形成 pzUSHIrK(TGC22) 载体 (SEQ. ID. NO. 41)。对 PCR 生成的克隆测序以确定不存在任何的 DNA 突变。

[0085] 根据本发明的方面，也可以将一种人干扰素 γ (hIFNg) 编码序列整合进一个质粒载体，也如图 8 中所图示显示的。所述 hIFNg 基因盒通过 PCR 扩增产生，具有侧翼的 BamHI 限制性内切酶的侧位点（引物序列见表 I）。然后将所述 BamHI 消化的 hIFNg 基因盒与原来被 eGFP 占据的 pzUSEIrK(TGC18) 的 BamHI 位点连接，形成 pzUSIfgIrK(TGC23) 载体 (SEQ. ID. NO. 37)。对 PCR 生成的克隆测序以确定不存在任何的 DNA 突变。

[0086] 表 I 用于产生其他表达基因的引物序列表

[0087]

引物名称	引物序列	序列I.D.
凝血酶原F	AAACCATGGCGCACGTCCGAGGC	SEQ. ID. NO. 1
凝血酶原R	CTACTCTCCAAACTGATCAATGA	SEQ. ID. NO. 2
皮离蛋白F	GGGGGATCCACCATGAGGTTCATGACTCTC	SEQ. ID. NO. 3
皮离蛋白R	GGCGGATCCCTATAGTACTGAGTCAAGG	SEQ. ID. NO. 4
hghF	GGGGGATCCACCATGGCTACAGGCTCCCGG	SEQ. ID. NO. 5
hghR	GGCGGATCCCTAGAAGCCACAGCTGCCC	SEQ. ID. NO. 6
hifngF	GGGGGATCCACCATGAAATATACAAGTTAT	SEQ. ID. NO. 7
hifngR	TCCGGATCCTTAATAAATAGATTTAGA	SEQ. ID. NO. 8
IFN pst-bam	CAACTGCAGGATCCAACAATGGCCTTGACCTTTGCTTTAC	SEQ. ID. NO. 9
IFN sac-xho	CAACTCGAGCTCATTCCTTACTTCTAAACTTTCTTG	SEQ. ID. NO. 10
eGFP bam [无起始密码子]	ATAGGATCCGTGAGCAAGGGCGAGGAGCTGTTC	SEQ. ID. NO. 11
eGFP sac [无起始密码子]	TATGAGCTCTTACTTGTACAGCTCGTCCATGCC	SEQ. ID. NO. 12
eGFP f bam [无起终止码子]	CAGGGATCCACCATGGTGAGCAAGGGCGAGG	SEQ. ID. NO. 13
eGFP r xho [无起终止码子]	AGTCTCGAGCTTGTACAGCTCGTCCATGC	SEQ. ID. NO. 14

[0088]

信号序列 AA pst bam	TATCTGCAGACCATGGTGAACAAACACTT CTTGTCCCTTTCGGTCCTCATCGTCCTCCT TGGCCTCTCCTCCAACTTGACAGCCGGGG GATCC	SEQ. ID. NO. 15
信号序列 AA ss-pcr	GTGAATTCGGATCCCCCCGGCTGTCAA	SEQ. ID. NO. 16
IRES sac	ATTGAGCTCAAGCTTCGAATTCTGCAG	SEQ. ID. NO. 17
IRES xho	ATACTCGAGGTGGCCATATTATCATCGTG TTTTTC	SEQ. ID. NO. 18
kan xho	ATACTCGAGACCATGATTGAACAAGATG GATTGCAC	SEQ. ID. NO. 19
Kan xba	ATTTCTAGACCAGAGCCGCCGCCAGCATT GACAGG	SEQ. ID. NO. 20
凝血酶原F	AAACCATGGCGCACGTCCGAGGC	SEQ. ID. NO. 21
凝血酶原R	CTACTCTCCAAACTGATCAATGA	SEQ. ID. NO. 22
皮离蛋白F	GGGGGATCCACCATGAGGTTCATGACTCT C	SEQ. ID. NO. 23
皮离蛋白R	GGCGGATCCCTATAGTACTGAGTCAAGG	SEQ. ID. NO. 24
hhghF	GGGGGATCCACCATGGCTACAGGCTCCC GG	SEQ. ID. NO. 25
hghR	GGCGGATCCCTAGAAGCCACAGCTGCCC	SEQ. ID. NO. 26
hifngF	GGGGGATCCACCATGAAATATACAAGTT AT	SEQ. ID. NO. 27
hifngR	TCCGGATCCTTAATAAATAGATTTAGA	SEQ. ID. NO. 28

[0089] 这些序列可以包含它们自己的信号序列。与被证明可以在芦荟植物中起作用的 α 干扰素中的序列类似，在凝血酶原、皮离蛋白 (DCD)、人生长激素 (hGH) 和人干扰素 γ (hIFNg) 天然信号序列也可以被证明在芦荟植物中发挥功能。无论如何，这些蛋白编码序列可以被克隆到一个包含 α 淀粉酶信号序列的载体中，如图 8 的显示的实例。

[0090] 在开发上述参考构建体的过程中，用于转化的所有载体在含 50ug/L 的博来霉素的 250ml LB 中生长，并且使用 CsCl 方法分离。然后，所述载体被应用于瞬时转染研究中，以确保每种构建体适当表达。使用基因枪进行玉米、土豆和芦荟的转染，并且对 eGFP 表达进行可视化监视或通过 rt-PCR 监测 IFN 或 kan 的表达。

[0091] 转化

[0092] 在生成包含目标基因和整合进载体的功能性的各种功能元件的 DNA 构建体之后，可以使用多种技术将所述 DNA 构建体引入到所述芦荟细胞中，所述技术是本领域技术人员通过阅读本公开内容后可以知晓的。通常对这些 DNA 构建体进行设计以促进稳定转化的芦荟植物 10 的形成。本领域中已知许多用重组 DNA 转化植物细胞的方法，所述方法可以被应用到本发明中。用于将包含在载体中的 DNA 构建体整合进芦荟细胞或组织以形成稳定的芦荟转化体的一些方法可以包括用根癌农杆菌或毛根农杆菌感染、用复制缺陷病毒感染、基因枪转化 (biolistic transformation)、原生质体转化、基因转移至花粉中、注射至生殖器官中、注射至未成熟胚中或者类似的方法。现在，两种最常用于植物转化的方法是农杆菌介导的转化和基因枪转化。转化的芦荟细胞或胼胝体组织通常在合适的营养培养基中生长，以筛选转化的细胞。经常地，选择培养基包含一种毒素或其他杀死非转化的细胞的选择因

子。培养基的各种改变将会产生包含目标基因的芦荟植物 10，这是本领域的技术人员通过阅读本公开内容可以知晓的。

[0093] 在转化的实施中，任何一次转化实验通常将 DNA 引入仅少部分的靶植物细胞中。标记基因被应用于为细胞的鉴定提供一个有效的系统，所述被鉴定的细胞是通过将转基因 DNA 构建体接受和整合进其基因组中以获得稳定的转化。优选的标记基因提供选择性标记，所述标记提供对一种选择剂（例如抗生素或除草剂）的抗性。对于选择性标记来说，被本发明的植物抵抗的任何除草剂都是有用的试剂。潜在的转化的细胞被暴露于所述选择剂。存活的细胞群通常将是整合赋予抗性的基因、并且以足够水平表达以使细胞存活的那些细胞。可以进一步地检测细胞以确定所述外源 DNA 的稳定整合。通常使用的选择性标记基因包括那些赋予抗生素抗性的基因，所述抗生素例如卡那霉素和巴龙霉素 (nptII)、潮霉素 B(aph IV) 和庆大霉素 (aac3 和 aacC4)；或者赋予除草剂抗性的基因，所述除草剂例如草胺磷 (bar orpat) 和草甘膦 (aroA 或 EPSPS)。这些选择的实例在编号为 5,550,318、5,633,435、5,780,708 和 6,118,047 的美国专利中有说明，所有这些专利通过引用的方式纳入本文。也可以使用能够用于可视化鉴定转化体的选择性标记，例如一种表达有色蛋白或萦光蛋白（例如萤光素酶或绿色萦光蛋白 (GFP)）的基因，或者一种表达 β 葡萄糖苷酸酶的基因，或者其各种显色底物已知的 uidA 基因 (GUS)。

[0094] 通常，在一种靶植物系的基因组中随机地引入重组 DNA（即在非特异性位点）是有用的。在特定的情况下，靶向重组 DNA 插入以获得位点特异性的整合可能是有用的，例如替换基因组中存在的基因、使用所述植物基因组中存在的启动子或者在已知会活化基因表达的预定位点插入一段重组多核苷酸。存在几种已知是功能性插入的特异性的重组系统，所述重组系统包括编号为 4,959,317 的美国专利公开的 cre-lox 和编号为 5,527,695 的美国专利公开的 FLP-FRT，所述两个专利通过引用的方式纳入本文并且在下文中更详细的叙述。

[0095] 对于基于根癌农杆菌的植物转化系统，存在于转化构建体上的其他元件将包括 T-DNA 左侧和右侧序列，以有利于将所述重组体多核苷酸整合进入所述植物基因组中。用农杆菌（根癌农杆菌）侵染，在室温下将芦荟的胼胝体组织与农杆菌的过夜培养物（包含整合有所述 DNA 构建体的载体）共孵育 1 个小时。所述农杆菌可以在合适的选择性培养基中生长。所述选择性培养基通常包含链霉素和卡那霉素。所述选择性培养基还可以包含乙酰丁香酮。浓度为 50uM-250uM 的乙酰丁香酮通常增加单子叶植物侵染性的效率。在所述选择性培养基上生长之后，可以将侵染的芦荟组织转移到培养皿的无选择性培养基中，在摄氏 25 度下黑暗中维持两 (2) 天。一种合适的示例性培养基可以是含有乙酰丁香酮的 MS 培养基。然后可以加入头孢噻肟 (cefotaxime)，维持两 (2) 天以杀死农杆菌。然后为细胞更换含有 50mg/L 的卡那霉素（用于筛选转化的芦荟细胞）和枝诱导生长因子 (0.2mg/L NAA, 2mg/L BAP) 的培养基，维持四 (4) 周，16 小时光照。然后可以将再生的枝转移至生根培养基中（含 0.2mg/L NAA 的 1/2MS) 以生成根（约 4-6 周后），之后转移至土壤。

[0096] 利用基因枪转化，直接用所述载体构建体轰击芦荟细胞或胼胝体组织。基因枪转化的各种实施方案和方面在编号为 5,015,580（大豆）、5,550,318（玉米）、5,538,880（玉米）、5,914,451（大豆）、6,160,208（玉米）、6,399,861（玉米）和 6,153,812（小麦）的美国专利中公开，并且农杆菌介导的转化在编号为 5,159,135（棉花）、5,824,877（大豆）、5,591,616（玉米）和 6,384,301（大豆）的美国专利中描述，所有这些专利通过引用的

方式纳入本文。此外，可以使用各种基因枪转化的装置，例如 Bio-Rad, Inc 的 Biolistic PDS-1000/He 粒子传递系统。在这种方法中，所述载体和关联的 DNA 构建体可以被绑定在金或其他颗粒上，并且被直接射进所述芦荟细胞。所述被轰击的芦荟细胞在黑暗中无选择性地生长 4 天，之后被转移到选择性的培养基。同样，所述选择性培养基可以包含 50mg/L 的卡那霉素。通常在 8-12 周内开始形成转化体，并且可以将其转移到长枝和生根培养基继续维持 8-12 周。一旦开始形成根，可以将小植物转移至土壤。

[0097] 也可以使用各种其他技术实现转化。这些技术包括使用复制妥协的病毒载体系统 (replication compromised viral vector system) 的病毒感染、电穿孔 (特别是对胼胝体细胞或生长在液体培养中的原生质体)，以及 PEG 或脂质介导的原生质体转移。然后可以按照与基因枪转化的步骤相同的基本步骤进行转化体的选择。

[0098] 一旦被引入所述芦荟胼胝体组织中，构建体可以被整合进入所述植物的遗传物质中，如图 3 中一般性显示的。在其他方面，所述构建体可以被稳定地整合到所述芦荟植物 10 的遗传物质外的细胞中。根据所述构建体，如果没有某种形式的转录诱导中，目标蛋白可以表达也可以不表达。在一方面，构建体一旦被引入芦荟细胞，就可以指导蛋白的合成或者指导运输至所述芦荟 10 的特定组织。典型地，这在目标蛋白中包含靶向信号序列的时候发生。

[0099] 位点特异性的整合 (Cre-Lox)

[0100] 在一方面，本发明可以利用位点特异性的整合或切出引入芦荟植物中的 DNA 构建体。位点特异性的整合或切出的一个优点是它可以被用来克服传统的转化技术涉及到的问题，其中转化构建体通常以多拷贝的形式随机地整合进一个宿主基因组中。如果所述外源 DNA 插入到一个必需基因中，这种将所要引入的 DNA 随机插入所述宿主细胞的基因组中可能是致死的。另外，一个转基因的表达可能被“位置效应”影响，所述位置效应是由周围基因组 DNA 引起。此外，因为与植物处理多转基因拷贝相关的困难，所述困难包括基因沉默、重组和不可预料的遗传，所以通常需要控制插入到所述转基因的芦荟植物 10 基因组中的 DNA 构建体的拷贝数量，并且经常仅需要插入单拷贝的所述 DNA 构建体。

[0101] 通过同源重组可以在植物中完成转基因或部分的转基因的位点特异性的整合或切出（例如见编号为 5,527,695 的美国专利，所述专利的全文通过引用的方式明确地纳入本文）。同源重组是任何具有相似核苷酸序列的 DNA 序列对之间的反应，其中所述两个序列相互作用（重组）以形成新的重组 DNA 片段。随着所述共享的核苷酸 DNA 序列长度的增加同源重组的频率增加，并且线性化的质粒分子的同源重组的频率比环形质粒分子高。在两个不完全相同的 DNA 序列之间可以出现同源重组，但是随着两个序列间差异的增加所述重组频率会降低。

[0102] 通过将目标 DNA 分子和与所述宿主细胞的内源序列有同源性的序列连接，通过同源重组可以定向引入的 DNA 序列。一旦所述 DNA 进入所述芦荟细胞，所述的两个同源序列可以相互作用，以在所述同源的基因组 DNA 序列所处位置插入所述引入的 DNA 序列。因此，对所述引入的 DNA 上包含的同源序列的选择将决定所述引入的 DNA 通过同源重组整合的位点。例如，如果目标 DNA 序列和与一个宿主芦荟细胞的单拷贝基因同源的 DNA 序列连接，那么目标 DNA 序列将通过同源重组仅插入到单个特异性位点。然而，如果目标 DNA 序列和与所述宿主真核细胞的多拷贝基因同源的 DNA 序列连接，那么目标 DNA 序列可以通过同源重

组在每个所述基因拷贝存在的特异性位点进行插入。

[0103] 通过涉及单互惠重组（导致插入全长的引入 DNA）或者通过双互惠重组（导致仅仅插入位于所述两个重组事件之间的 DNA）的同源重组反应，可以在所述宿主基因组中插入 DNA。例如，如果希望将一个外源基因插入到一个选定基因所处的基因组位点中，所述引入的 DNA 应该包含与所述选定的基因同源的序列。然后单个的重组事件会导致所述整个的引入 DNA 序列被插入到所选定的基因中。或者，通过将目标 DNA 序列（所述意欲插入到所述基因组中的序列）的每一端侧接与所选定的基因同源的 DNA 序列，可以实现双重组事件。涉及每个所述同源侧翼区的同源重组事件将导致所述外源 DNA 的插入。因此，仅仅那些位于共享基因组同源性的两个区域之间的 DNA 序列可以被整合进入所述基因组中。

[0104] 虽然引入的 DNA 序列可以通过同源重组被定向插入到特异性基因组位点，但是在高等真核生物中，与随机插入事件相比，同源重组是相对少的事件。在芦荟细胞中，外源 DNA 分子在所述芦荟植物的基因组中发现同源序列，并且以大约 $0.5\text{-}4.2\times10^{-4}$ 的频率重组。因此，包含通过同源重组整合的引入 DNA 序列的任何转化的细胞将也可能包含许多拷贝的随机整合引入的 DNA 序列。避免这些随机插入的一种方法是使用一种位点特异性重组酶系统。通常，一种位点特异性重组酶系统由三个元件组成：两对 DNA 序列（所述位点特异性重组序列）和一种特异性酶（所述位点特异性重组酶）。所述位点特异性重组酶将仅催化两个位点特异性重组序列之间的重组反应。

[0105] 根据本发明，可以使用大量的不同的位点特异性重组酶系统，包括但不限于噬菌体 P1 的 Cre/lox 系统（编号为 5,658,772 的美国专利，所述专利的全文通过引用的方式明确地纳入本文）、酵母的 FLP/FRT 系统 (Golic and Lindquist, 1989)、噬菌体 Mu 的 Gin 重组酶 (Maeser and Kahmann, 1991)、E. coli 的 Pin 重组酶 (Enomoto et al., 1983) 和 pSR1 质粒的 R/RS 系统 (Araki et al., 1992)。所述噬菌体 P1 Cre/lox 和酵母 FLP/FRT 系统组成用于转基因的位点特异性整合或切出的两种特别有用的系统。在这些系统中，一种重组酶 (Cre 或 FLP) 将特异性地与其各自的位点特异性重组序列（分别是 lox 或 FRT）相互作用，以反转或切出所述间隔序列。这两个系统的序列相对比较短 (lox 为 34bp 并且 FRT 为 47bp)，并且因此方便与转化载体联合使用。

[0106] 图 8 显示了使用 Flp/Frt 和 Cre/Loxp 重组酶系统的一种具体的示例性方案。在一个这样的系统中，来自噬菌体和酵母的位点特异性重组酶可以被用作在试管中和在活体器官中操作 DNA 的工具。重组酶 / 重组位点的组合可以包括 Cre-Lox、FLP-FRT 和 R-RS，其中 Cre、FLP 和 R 是重组酶，并且 Lox、FRT 和 RS 是重组位点。为了使用这个系统，可以产生一种包含用于重组的特异性位点的转基因植物。通过转染一种载体、表达一种选择性标记并且设计包含串联排列的 Frt 和 Lox 位点，产生一种亲代的转基因的芦荟植物 10。所述 Frt 和 Lox 位点都由三个元件组成，在每个长度都是十三个 (13) 核苷酸的两个 (2) 反向重复之间的长度为八个 (8) 核苷酸的间隔序利。所述反向重复序列作为所述特异性重组酶的 DNA 结合结构域，而所述间隔元件是可变的，但是对于同源重组是必须的。通过改变 LoxP 或者 FRT 位点处的间隔元件，可以产生用于重组的多个不同位点。通过改变 Frt 和 Lox 位点，可以形成一个允许多位点定向整合的系统（如在图 8 中列出的）。该系统可以具有许多优点。生成所述的初始的亲代植物之后，使内源基因的破坏最小。通过位点特异性重组酶的表达可以提高所述转化的效率。使用串联排列的 lox 和 frt 位点使得可以选择性的除去选

择的标记而保留目标基因。并且一旦形成一株亲代植物，细胞的增殖和再生能力会得到提高。然而，所述亲代植物是第一步，并且按照标准的转化方法（基因枪或农杆菌）形成。通过一种包含 LoxP 和 Frt 位点以及目标基因和一种选择性标记的载体与一种表达所述 Cre 重组酶的载体共转染，形成后续的转基因植物。Cre 重组酶的表达诱导所述载体的 LoxP 位点和所述亲本植物的同源重组。通过表达选择性标记选择转化体，并且诱导所述转化体再生。用表达 Flp 重组酶的后续转化可以除去所述选择性标记并且使得可以随后整合到其他 Loxp 位点中。

[0107] 产生存活的植物

[0108] 在芦荟细胞或芦荟组织转化之后，所述转化的芦荟细胞或芦荟组织可以生长成小植物。可以在再生培养基中培养在所述选择剂暴露中存活的芦荟细胞或者在筛选实验中得分为正的芦荟细胞，并且使它们发育成转基因的芦荟植物 10。可以将由转化的芦荟细胞再生的发育中的芦荟小植物转移到植物生长基质并硬化，例如在一个大约 85％的相对湿度、600ppm CO_2和 25-250 微爱因斯坦 (microeinstein)m $^{2}S^{1}$的光照的环境控制箱中，然后转移至一个温室或生长室中成熟。依赖于所述初始组织，在转化体被鉴定后，所述转化的芦荟植物 10 再生大约 6 周至 10 个月。通常可以测试所述再生的转化芦荟植物 10 或其子代种子或植物的所述重组 DNA 的表达。

[0109] 从转化的未分化的胼胝体组织或分生组织再生的转基因芦荟植物 10 主要涉及生长因子生长素和激动素的处理。所述第一步是启动枝的生长。枝诱导培养基通常包含低浓度的生长素 (0.2mg/L NAA) 和相对高浓度的细胞分裂素 (2mg/L BAP)。降低糖浓度至 2％的蔗糖（影响渗透压）也可有助于植物的再生。在生枝培养基上的细胞被置于 25℃的培养箱中，以 12 小时的白昼循环。在开始形成的枝达到约 2cm 的长度后将其从所述原始胼胝体上切除，并且置于生根培养基中。也可以将这些根维持在生枝培养基中以促进其他枝的发育。生根培养基促进根的发育，并且通常包含 1/2 的 MS 培养基和相对低浓度的生长素 (0.2mg/L NAA)。在根开始发育之后 (1-2cm)，将所述小植物转移至土壤众并且逐渐减少所述相对湿度使之适应环境。一旦产生转基因的芦荟植物 10 之后，可以通过种子、克隆繁殖的方法或其他的方法繁殖所述转基因的芦荟植物 10，所述其他方法是本领域的技术人员通过阅读本公开内容后可以知晓的。

[0110] 从植物或植物细胞提取并纯化蛋白

[0111] 通过载体引入到植物细胞或植物中的所述结构的蛋白或酶可以存在于所述芦荟植物 10 的各种组织中，所述组织包括根、块茎、叶、种子、花、树液或植物部分和发育阶段的结合。在一方面，目标一种或多种蛋白可以富集在所述叶的中央部分的凝胶基质中。在另一方面，当从所述芦荟叶 12 中提取之后，一种或多种目标蛋白可以是生物活性的并且可以提供某种程度的药用效果。在另一方面，一种或多种目标蛋白与在所述凝胶基质中存在的至少一种自身的芦荟蛋白、碳水化合物和 / 或其他化合物协同地作用，以便为患者提供所需的治疗。可以从总生物量或个别的组织提取蛋白。从植物细胞提取蛋白的方法可以包括物理的和化学的方法。从叶中或树液中提取蛋白的方法可以涉及到过滤、超速离心、化学萃取和亲和层析。

[0112] 在一方面，目标蛋白可能积聚在所述转化的芦荟植物 10 的叶的凝胶中。这个区域主要是水，大约 99％，并且没有有害的蛋白酶。所述凝胶的提取是本领域的技术人员可以理

解的已知技术。一些蛋白，特别是化妆品用的蛋白，可以直接被用于补充所述芦荟凝胶，并且绝不需要从所述凝胶中将其单独地分离出来。

[0113] 实施例

[0114] 分离植物细胞的方法

[0115] 在第一个实例中，用于直接应用或用于胼胝体的初始化的枝分生组织分离自所述芦荟的茎，所述芦荟是芦荟 (Aloe vera)（库拉索芦荟）、开普芦荟 (Aloe ferox) 或木立芦荟 (Aloe arborescence)。去除叶部分，并且沿所述芦荟叶的纵轴方向从侧面将茎切开。用吐温 20(0.05％）和次氯酸盐 (5％）的混合液将组织灭菌 10 分钟，之后用乙醇浸泡 30 秒钟，用 3x 的无菌水冲洗。然后将切片暴露表面侧向放置在含琼脂 (0.8-1％）的 MS 培养基的平皿中，所述 MS 培养基包含各种浓度的所述生长因子、生长素和细胞分裂素，以及 2-3％的蔗糖。根据所需的过程，改变生长条件。无枝发育的胼胝体培养通常是通过在没有 BAP 或含有低浓度的 BAP(0.2mg/L) 的 NAA(2-5mg/L) 中培养这些切片开始。未分化的细胞开始沿被切割的茎的区域生长，并且可以在不同的皿上传代和维持。通过将这种片段置于生枝培养基上可以直接开始枝诱导，所述生枝培养基是包含 0.2mg/L NAA（或 IAA 0.2mg/L) 和 2mg/L BAP 的 MS 培养基。NAA 和 IAA 二者都可以作为生长素的来源，并且有效果的生长素和细胞分裂素都有一个浓度范围。来自未成熟胚的胼胝体细胞可以从市售的种子发育而来。所述种子首先被灭菌（乙醇 30 秒，5％次氯酸盐加 0.05％的吐温 20 中 15 分钟，并在无菌水中清洗 3 次）。然后将灭菌的种子置于水中，4 摄氏度下过夜。通过除去种皮分离未成熟胚。在所述三角形的种子的一个角切开一个小口，并且将胚挤出。然后将它置于 MS 培养基上，所述培养基仅包含生长素 (NAA 2-5mg/L) 以诱导胼胝体，或者包含生长素和细胞分裂素 (NAA 0.2mg/L,BAP 2mg/L) 以诱导胼胝体和枝增殖。糖浓度通常是 2-3％。

[0116] 在去叶的年轻芦荟植物 10 的底部剪下的 1-2 厘米的片段，由所述片段可以发育成来自所述枝顶端分生组织的胼胝体细胞。所述片段经过表面灭菌，重新放置于含 1％琼脂的 MS 培养基中，所述培养基含生长素和细胞分裂素 (2mg/L NAA、0.2mg/L BAP)。

[0117] 在 1-2 周内开始胼胝体细胞的诱导。然后可以从这些培养物中分离细胞，并且将所述细胞用作转化的起始材料。

[0118] DNA 构建体

[0119] i. pBI121 泛素干扰素载体 (pBI-UI)。

[0120] 所述 pBI121 农杆菌二元载体被用作产生 pBI-UI 的骨架。使用包含 5’PstI 和 3’SacI/XhoI 限制性位点的基因特异性引物从人 293 个细胞中扩增具有信号序列的人干扰素 α-2 基因。所述 587bp 的干扰素 PCR 产物被克隆到 pZErO 中并测序。通过扩增来自所述载体 pUBI-GFP 的区域并克隆进 5’Hind Ⅲ和 3’PstI 限制性位点中，来自玉米的 1962bp 的泛素 (Ubi) 启动子元件也被克隆进 pZErO。然后，使用 Hind Ⅲ /PstI(pZErO UI) 将所述 Ubi 片段亚克隆进入所述 pZero 干扰素载体。然后使用 Hind Ⅲ /PstI 将所述完整的 Ubi 干扰素盒亚克隆到 pBI121 中，并且去除所述 CaMV 35S 启动子和 GUS 基因形成 pBIUI。该载体保留了农杆菌介导的感染和转化必需的右侧和左侧 T- 元件边缘区域，并且转化和表达受 Nos 启动子控制的卡那霉素抗性基因 (NPTII)。

[0121] ii. pZErO 泛素干扰素 IRES 卡那霉素。(pZ UIIK)

[0122] 该载体的产生是通过将所述 Ubi 启动子克隆进所述 pZErO 克隆载体。在该片段的

之后克隆一个盒，所述盒包含人干扰素 α-2、IRES（内部核糖体进入序列）和卡那霉素抗性基因。该单元被表达为单个的转录物，但被翻译成两种不同的蛋白（由 IRES 提供的第二翻译起始位点导致）。这使得所述选择性标记和干扰素都在所述 Ubi 启动子的控制下表达。

[0123] iii. 使用 Cre 和 Flp 重组酶的双质粒系统。

[0124] 首先产生一个亲代植物，所述植物可表达具有用于定向整合的 cre 和 flp 位点的选择性标记。引入具有目标基因的载体的另一次转染被靶向所述 cre 位点。然后使用 flp 去除不需要的遗传物质。

[0125] 将 DNA 构建体导入分离的植物细胞

[0126] i. 农杆菌介导的基因转移

[0127] 用所述 pBIUI 载体电穿孔农杆菌菌株 LB4404(Invitrogen,Carlsbad,CA)，并且在含 50ug/ml 的卡那霉素和 100ug/ml 链霉素的 LB 琼脂平板上筛选阳性克隆。摄氏 30 度下，使单克隆在含 250uM 的乙酰丁香酮的 LB 培养基中过夜生长。感染的发生是通过将植物片段或胼胝体浸入所述过夜的培养物中，在无菌滤纸上干燥印迹，并且在摄氏 25 度于黑暗中放置于包含 250uM 的乙酰丁香酮的 MS 琼脂平板上。感染两天之后，将组织转移到包含 200uM 的头孢噻肟 (cefotaxime) 的 MS 琼脂平板上以杀死农杆菌。然后将所述组织转移到含 50mg/L 的卡那霉素、0.2mg/L 的 NAA 和 2mg/L 的 BAP 的 MS 琼脂平板上以筛选转化体，并且诱导枝再生。所述组织在 12 小时光照和摄氏 25 度下生长。使不定枝生长到大约 2cm 的长度，然后将其剪下并且移植到生根培养基上继续筛选。检测表达目标基因的小植物的表达水平，并且将其转移到土壤中。

[0128] ii. 基因枪基因转移

[0129] 用包含所述 DNA 构建体的载体包被金颗粒 (0.6-1um)。所述金颗粒在 70％的乙醇中搅拌并浸泡清洗 15 分钟，之后在无菌水中清洗 3 次。然后将所述金颗粒重新悬浮于 50％的丙三醇中至 60mg/ml 的浓度。为了在所述清洗的颗粒上包被载体 DNA，将 3mg 的颗粒加入 1.5ml 的微量离心管中。在所述微量离心管中加入 5ul 浓度为 1ug/ul 的 DNA、50ul 2.5M $CaCl_2$ 和 20ul 0.1M 亚精胺，并且搅拌 2-3 分钟。使之沉降、离心 1-2 秒并弃去液体。加入 140ul 70％的乙醇，离心并弃去液体。加入 140ul 100％的乙醇，离心并弃去液体。在沉淀中加入 48ul 100％的乙醇并轻轻地再悬浮。将所述基因枪装置 (PDS 1000/He,Bio-Rad Laboratories,Inc.,Hercules,CA) 用 70％的乙醇灭菌。使用巨携带子支撑物 (macrocarrier holder) 和停止屏（推荐设置）之间的最小间隙距离装配所述基因枪装置。用移液管将包被的金颗粒 (8ul) 滴入所述巨携带子中央。根据所述靶组织 (6cm 组织，9cm 胼胝体）设定所述靶距离，破圆盘 (rupture disk) 为 600-1100psi。

[0130] 在第一实例中，在轰击后细胞被置于包含 2mg/L NAA 的无选择性 MS 琼脂平板上，在摄氏 25 度下的黑暗中培养 1 周。然后将细胞转移到含用于选择的 50mg/L 的卡那霉素的 MS 琼脂平板上继续 4-5 周，所述琼脂平板含生长素 (2mg/L NAA) 和细胞分裂素 (0.2mg/L,6- 苄氨基嘌呤）。所述细胞在摄氏 25 度下以 12 小时的光照生长。

[0131] 在第二实例中，在一个 MS 的平板 (4g/L MS 盐、1mg/L(1000X)MS 维生素储液、2mg/L NAA、100mg/L 肌醇、2.76g/L 脯氨酸、30g/L 蔗糖、100mg/L 酪蛋白水解产物、36.4g/L 山梨糖醇、36.4g/L 甘露醇、2.5g/L 固化剂 (gelrite)，pH 5.8) 上进行微粒轰击。在轰击后，所述轰击的胼胝体留在 MSOSM 上 1 个小时，然后转移到 MS 初始培养基中 (4g/L MS 盐、1mg/

L(1000X)MS 维生素储液、2mg/L NAA、100mg/L 肌醇、2g/L 脯氨酸、30g/L 蔗糖、100mg/L 酪蛋白水解产物、2.5g/L 固化剂 (gelrite),pH 5.8)。在 MS 上 7-10 天之后,所述轰击的胚胎体被转移到含 50mg/L 的卡那霉素的 MSS 选择培养基中 (4g/L MS 盐、1mg/L(1000X)MS 维生素储液、2mg/L NAA、100mg/L 肌醇、30g/L 蔗糖、2.5g/L 固化剂 (gelrite),pH 5.8) 用于筛选转化的细胞。在 3 周后,单胚胎体碎片被转移到新鲜 MSS 培养基中。在轰击的 6-8 周内,由选定的胚胎体碎片产生卡那霉素抗性克隆。

[0132] iii. 原生质体

[0133] 原生质体是没有其外细胞壁的植物细胞。产生这些细胞的优点是提高转染的效率,并且可以将这样的细胞融合形成一种体细胞杂种。体细胞杂种可以将来自不同植物的基因混合,有可能获得完全新的细胞。

[0134] 使用根据本发明的原生质体的技术可以包括在液体培养中生长所述细胞 (对数生长期中的 2-3 天)。然后离心所述细胞 (1000x 重力,5 分钟)。所述细胞被重新悬浮于含纤维素酶、浸解酶和果胶酶的溶液中。在摄氏 25 度黑暗中振荡过夜。离心 (600x 重力,5 分钟) 并除去上清液。在培养基中重新悬浮所述原生质体,并且多次离心以除去微量的酶。所述原生质体不能被存储或者用于增殖,并且必须被尽快用于转染或体细胞融合。

[0135] 产生包含 DNA 构建体的存活的植物

[0136] 在第一个实例中,利用其细胞毒素 (例如卡那霉素) 的抗性筛选成功的转化体。稳定转染的细胞必须表达所述抗性标记和目标基因,例如编码干扰素的基因,以及能够再生成完整的植物。将 50mg/L 的卡那霉素加到所述 MS 琼脂平板以启动筛选。所述筛选过程部分地依赖于所述细胞复制的速度,但是可以在 6-10 周间发生。在这个时间中,通过在所述 MS 琼脂平板中加入 0.2mg/L NAA 和 2mg/L BAP 诱导转化的组织的再生,并且用 12 小时 / 光照循环促进枝的发育 (继续使用选择压力)。当不定枝开始发育的时候,在大约 2cm 长时可以对它们进行基因表达的检测。随后将表达目标基因产物的枝 (利用 RT-PCR、Western 印迹和生物检验测定) 转移到 MS 琼脂平板上以诱导根形成 (1/2MS 加 0.2mg/L NAA)。

[0137] 在第二个实例中,转基因胚胎体的再生的实现是通过将大约 15 个小碎片 (约 4mm) 转移到含有经过滤灭菌的卡那霉素 (50mg/L) 的再生培养基 I 中 (4.3g/L MS 盐、1ml/L(1000X)MS 维生素储液、100mg/L 肌醇、60g/L 蔗糖、3g/L 固化剂 (gelrite),pH 5.8),并且在摄氏 25 度下黑暗中孵育 2-3 周。在光照下将成熟的体细胞胚转移至在含有经过滤灭菌的卡那霉素 (50mg/L) 的再生培养基 II 中 (4.3g/L MS 盐、1ml/L(1000X)MS 维生素储液、100mg/L 肌醇、30g/L 蔗糖、3g/L 固化剂 (gelrite),pH 5.8) 以进行萌发。

[0001]

序列表

<110> 绿色细胞有限公司
W.J.劳瑟
W-S.约
K.洛里克
M.佩奇斯
N.J.劳特

<120> 用于生产蛋白的转基因芦荟植物及其相关的方法

<130> 1015-003

<160> 44

<170> PatentIn version 3.3

<210> 1
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 1
aaaccatggc gcacgtccga ggc 23

<210> 2
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 2
ctactctcca aactgatcaa tga 23

<210> 3
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 3
gggggatcca ccatgaggtt catgactctc 30

<210> 4
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 4
ggcggatccc tatagtactg agtcaagg 28

[0002]

<210> 5
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 5
gggggatcca ccatggctac aggctcccgg 30

<210> 6
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 6
ggcggatccc tagaagccac agctgccc 28

<210> 7
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 7
gggggatcca ccatgaaata tacaagttat 30

<210> 8
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 8
tccggatcct taataaatag atttaga 27

<210> 9
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 9
caactgcagg atccaacaat ggccttgacc tttgctttac 40

<210> 10
<211> 36
<212> DNA
<213> 人工序列

[0003]

<220>
<223> 合成寡核苷酸

<400> 10
caactcgagc tcattcctta cttctaaact ttcttg 36

<210> 11
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 11
ataggatccg tgagcaaggg cgaggagctg ttc 33

<210> 12
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 12
tatgagctct tacttgtaca gctcgtccat gcc 33

<210> 13
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 13
cagggatcca ccatggtgag caagggcgag g 31

<210> 14
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 14
agtctcgagc ttgtacagct cgtccatgc 29

<210> 15
<211> 93
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 15
tatctgcaga ccatggtgaa caaacacttc ttgtcccttt cggtcctcat cgtcctcctt 60

[0004]

ggcctctcct ccaacttgac agccgggggga tcc 93

<210> 16
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 16
gtgaattcgg atcccccggc tgtcaa 26

<210> 17
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 17
attgagctca agcttcgaat tctgcag 27

<210> 18
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 18
atactcgagg tggccatatt atcatcgtgt ttttc 35

<210> 19
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 19
atactcgaga ccatgattga acaagatgga ttgcac 36

<210> 20
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 20
atttctagac cagagccgcc gccagcattg acagg 35

<210> 21

[0005]

<210> 23
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 21
aaaccatggc gcacgtccga ggc 23

<210> 22
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 22
ctactctcca aactgatcaa tga 23

<210> 23
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 23
gggggatcca ccatgaggtt catgactctc 30

<210> 24
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 24
ggcggatccc tatagtactg agtcaagg 28

<210> 25
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 25
gggggatcca ccatggctac aggctcccgg 30

<210> 26
<211> 28
<212> DNA
<213> 人工序列

<220>

[0006]

<223> 合成寡核苷酸

<400> 26
ggcggatccc tagaagccac agctgccc 28

<210> 27
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 27
gggggatcca ccatgaaata tacaagttat 30

<210> 28
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 28
tccggatcct taataaatag atttaga 27

<210> 29
<211> 90
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 29
ctgcagacca tggtgaacaa acacttcttg tcccttttcgg tcctcatcgt cctccttggc 60

ctctcctcca acttgacagc cgggggatcc 90

<210> 30
<211> 333
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 30
atgaggttca tgactctcct cttcctgaca gctctggcag gagccctggt ctgtgcctat 60

gatccagagg ccgcctctgc cccaggatcg gggaaccctt gccatgaagc atcagcagct 120

caaaaggaaa atgcaggtga agacccaggg ttagccagac aggcaccaaa gccaaggaag 180

cagagatcca gccttctgga aaaaggccta gacggagcaa aaaaagctgt gggggggactc 240

ggaaaaactag gaaaagatgc agtcgaagat ctagaaagcg tgggtaaagg agccgtccat 300

gacgttaaag acgtccttga ctcagtacta tag 333

[0007]

<210> 31
<211> 720
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 31
atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac 60

ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120

ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc 180

ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga ccacatgaag 240

cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc 300

ttcaaggacg acggcaacta caagacccgc gccgaggtga agttcgaggg cgacaccctg 360

gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420

aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac 480

ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc 540

gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac 600

tacctgagca cccagtccgc cctgagcaaa gaccccaacg agaagcgcga tcacatggtc 660

ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct gtacaagtaa 720

<210> 32
<211> 669
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 32
ggatccacca tggctacagg ctcccggacg tccctgctcc tggcttttgg cctgctctgc 60

ctgccctggc ttcaagaggg cagtgccttc ccaaccattc ccttatccag gctttttgac 120

aacgctatgc tccgcgccca tcgtctgcac cagctggcct ttgacaccta ccaggagttt 180

gaagaagcct atatcccaaa ggaacagaag tattcattcc tgcagaaccc ccagacctcc 240

ctctgtttct cagagtctat tccgacaccc tccaacaggg aggaaacaca acagaaatcc 300

aacctagagc tgctccgcat ctccctgctg ctcatccagt cgtggctgga gcccgtgcag 360

ttcctcagga gtgtcttcgc caacagcctg gtgtacggcg cctctgacag caacgtctat 420

gacctcctaa aggacctaga ggaaggcatc caaacgctga tggggaggct ggaagatggc 480

agccccccgga ctgggcagat cttcaagcag acctacagca agttcgacac aaactcacac 540

aacgatgacg cactactcaa gaactacggg ctgctctact gcttcaggaa ggacatggac 600

aaggtcgaga cattcctgcg catcgtgcag tgccgctctg tggagggcag ctgtggcttc 660

[0008]

tagggatcc 669

<210> 33
<211> 570
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 33
acaatggcct tgacctttgc tttactggtg gccctcctgg tgctcagctg caagtcaagc 60

tgctctgtgg gctgtgatct gcctcaaacc cacagcctgg gtagcaggag gaccttgatg 120

ctcctggcac agatgaggag aatctctctt ttctcctgct tgaaggacag acatgacttt 180

ggatttcccc aggaggagtt tggcaaccag ttccaaaagg ctgaaaccat ccctgtcctc 240

catgagatga tccagcagat cttcaatctc ttcagcacaa aggactcatc tgctgcttgg 300

gatgagaccc tcctagacaa attctacact gaactctacc agcagctgaa tgacctggaa 360

gcctgtgtga tacaggggct gggggtgaca gagactcccc tgatgaagga ggactccatt 420

ctggctgtga ggaaatactt ccaaagaatc actctctatc tgaaagagaa gaaatacagc 480

ccttgtgcct gggaggttgt cagagcagaa atcatgagat ctttttcttt gtcaacaaac 540

ttgcaagaaa gtttaagaag taaggaatga 570

<210> 34
<211> 584
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 34
gcccctctcc ctccccccca cctaacgtta ctggccgaag ccgcttggaa taaggccggt 60

gtgcgtttgt ctatatgtta ttttccacca tattgccgtc ttttggcaat gtgagggccc 120

ggaaacctgg ccctgtcttc ttgacgagca ttcctagggg tctttccccc ctcgccaaag 180

gaatgcaagg tctgttgaat gtcgtgaagg aagcagttcc tctggaagct tcttgaagac 240

aaacaacgtc tgtagcgacc ctttgcaggc agcggaaccc cccacctggc gacaggtgcc 300

tctgcggcca aaagccacgt gtataagata cacctgcaaa ggcggcacaa ccccagtgcc 360

acgttgtgag ttggatagtt gtggaaagag tcaaatggct ctcctcaagc gtattcaaca 420

aggggctgaa ggatgcccag aaggtacccc attgtatggg atctgatctg gggcctcggt 480

gcacatgctt tacatgtgtt tagtcgaggt taaaaaaacg tctaggcccc ccgaaccacg 540

gggacgtggt tttcctttga aaaacacgat gataatatgg ccac 584

<210> 35

[0009]

<211> 1469
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 35
atgattgaac aagatggatt gcacgcaggt tctccggccg cttgggtgga gaggctattc 60

ggctatgact gggcacaaca gacaatcggc tgctctgatg ccgccgtgtt ccggctgtca 120

gcgcaggggc gcccggttct ttttgtcaag accgacctgt ccggtgccct gaatgaactg 180

caggacgagg cagcgcggct atcgtggctg gccacgacgg gcgttccttg cgcagctgtg 240

ctcgacgttg tcactgaagc gggaagggac tggctgctat tgggcgaagt gccggggcag 300

gatctcctgt catctcacct tgctcctgcc gagaaagtat ccatcatggc tgatgcaatg 360

cggcggctgc atacgcttga tccggctacc tgcccattcg accaccaagc gaaacatcgc 420

atcgagcgag cacgtactcg gatggaagcc ggtcttgtcg atcaggatga tctggacgaa 480

gagcatcagg ggctcgcgcc agccgaactg ttcgccaggc tcaaggcgcg catgcccgac 540

ggcgatgatc tcgtcgtgac ccatggcgat gcctgcttgc cgaatatcat ggtggaaaat 600

ggccgctttt ctggattcat cgactgtggc cggctgggtg tggcggaccg ctatcaggac 660

atagcgttgg ctacccgtga tattgctgaa gagcttggcg gcgaatgggc tgaccgcttc 720

ctcgtgcttt acggtatcgc cgctcccgat tcgcagcgca tcgccttcta tcgccttctt 780

gacgagttct tctgagcggg actctggggt tcgaaatgac cgaccaagcg acgcccaacc 840

tgccatcacg agatttcgat tccaccgccg ccttctatga aaggttgggc ttcggaatcg 900

ttttccggga cgccggctgg atgatcctcc agcgcgggga tctcatgctg gagttcttcg 960

cccacgggat ctctgcggaa caggcggtcg aaggtgccga tatcattacg acagcaacgg 1020

ccgacaagca caacgccacg atcctgagcg acaatatgat cgggcccggc gtccacatca 1080

acggcgtcgg cggcgactgc ccaggcaaga ccgagatgca ccgcgatatc ttgctgcgtt 1140

cggatatttt cgtggagttc ccgccacaga cccggatgat ccccgatcgt tcaaacattt 1200

ggcaataaag tttcttaaga ttgaatcctg ttgccggtct tgcgatgatt atcatataat 1260

ttctgttgaa ttacgttaag catgtaataa ttaacatgta atgcatgacg ttatttatga 1320

gatgggtttt tatgattaga gtcccgcaat tatacattta atacgcgata gaaaacaaaa 1380

tatagcgcgc aaactaggat aaattatcgc gcgcggtgtc atctatgtta ctagatcggg 1440

cctcctgtca atgctggcgg cggctctgg 1469

<210> 36
<211> 1874
<212> DNA
<213> 人工序列

<220>

[0010]

〈223〉 合成寡核苷酸

〈400〉 36

aaaccatggc gcacgtccga ggcttgcagc tgcctggctg cctggccctg gctgccctgt 60

gtagccttgt gcacagccag catgtgttcc tggctcctca gcaagcacgg tcgctgctcc 120

agcgggtccg gcgagccaac accttcttgg aggaggtgcg caagggcaac ctagagcgag 180

agtgcgtgga ggagacgtgc agctacgagg aggccttcga ggctctggag tcctccacgg 240

ctacggatgt gttctgggcc aagtacacag cttgtgagac agcgaggacg cctcgagata 300

agcttgctgc atgtctggaa ggtaactgtg ctgagggtct gggtacgaac taccgagggc 360

atgtgaacat cacccggtca ggcattgagt gccagctatg gaggagtcgc tacccacata 420

agcctgaaat caactccact acccatcctg gggccgacct acaggagaat ttctgccgca 480

accccgacag cagcaccacg ggaccctggt gctacactac agaccccacc gtgaggaggc 540

aggaatgcag catccctgtc tgtggccagg atcaagtcac tgtagcgatg actccacgct 600

ccgaaggctc cagtgtgaat ctgtcacctc cattggagca gtgtgtccct gatcgggggc 660

agcagtacca ggggcgcctg gcggtgacca cacatgggct cccctgcctg gcctgggcca 720

gcgcacaggc caaggccctg agcaagcacc aggacttcaa ctcagctgtg cagctggtgg 780

agaacttctg ccgcaaccca gacggggatg aggagggcgt gtggtgctat gtggccggga 840

agcctggcga ctttgggtac tgcgacctca actattgtga ggaggccgtg gaggaggaga 900

caggagatgg gctggatgag gactcagaca gggccatcga agggcgtacc gccaccagtg 960

agtaccagac tttcttcaat ccgaggacct ttggctcggg agaggcagac tgtgggctgc 1020

gacctctgtt cgagaagaag tcgctggagg acaaaaccga aagagagctc ctggaatcct 1080

acatcgacgg gcgcattgtg gagggctcgg atgcagagat cggcatgtca ccttggcagg 1140

tgatgctttt ccggaagagt ccccaggagc tgctgtgtgg ggccagcctc atcagtgacc 1200

gctgggtcct caccgccgcc cactgcctcc tgtacccgcc ctgggacaag aacttcaccg 1260

agaatgacct tctggtgcgc attggcaagc actcccgcac aaggtacgag cgaaacattg 1320

aaaagatatc catgttggaa aagatctaca tccacccccag gtacaactgg cgggagaacc 1380

tggaccggga cattgccctg atgaagctga agaagcctgt tgccttcagt gactacattc 1440

accctgtgtg tctgcccgac agggagacgg cagccagctt gctccaggct ggatacaagg 1500

ggcgggtgac aggctggggc aacctgaagg agacgtggac agccaacgtt ggtaaggggc 1560

agcccagtgt cctgcaggtg gtgaacctgc ccattgtgga gcggccggtc tgcaaggact 1620

ccacccggat ccgcatcact gacaacatgt tctgtgctgg ttacaagcct gatgaaggga 1680

aacgagggga tgcctgtgaa ggtgacagtg ggggaccctt tgtcatgaag agccccttta 1740

acaaccgctg gtatcaaatg ggcatcgtct catggggtga aggctgtgac cgggatggga 1800

aatatggctt ctacacacat gtgttccgcc tgaagaagtg gatacagaag gtcattgatc 1860

[0011]

agtttggaga gtag 1874

<210> 37
<211> 4688
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 37
aagctttgat cccctaatga gcattgcatg tctaagttat aaaaaattac cacatatttt 60
tttgtcacac ttgtttgaag tgcagtttat ctatctttat acatatattt aaacttcact 120
ctacgaataa tataatctat agtactacaa taatatcagt gttttagaga atcatataaa 180
tgaacagtta gacatggtct aaaggacaat tgagtacttt gacaacagga ctctacagtt 240
ttatctcttt agtgtgcatg tgttctcctt ttttttttgca aatagcttca cctatataat 300
acttcatcca ttttattagt acatccattt agggtttagg gttaatggtt tttatagact 360
aatttttttta gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact 420
ctattttagt ttttttattt aataatttag atataaaata gaataaaata aagtgactaa 480
aaattaaaca aatacccttt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg 540
agtagataat gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa 600
ccagcagcgt cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct 660
gggcccctct cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa 720
ttgcgtggcg gagcggcaga cgtgagccgg cacggcaggc ggcctcctcc tcctctcacg 780
gcaccggcag ctacgggggga ttcctttccc accgctcctt cgctttcccc tcctcgcccg 840
ccgtaataaa tagacacccc ctccacaccc tctttccccca acctcgtgtt gttcggagcg 900
cacacacaca caaccagatc tcccccaaat ccacccgtcg gcacctccgc ttcaaggtac 960
gccgctcgtc ctcccccccc ccctctctac cttctctaga tcggcgttcc ggtccatggt 1020
tagggcccgg tagttctact tctgttcatg tttgtgttag atccgtgttt gtgatagatc 1080
cgtgctgcta gcgttcgtac acggatgcga cctgtacgtc agacacgttc tgattgctaa 1140
cttgccagtg tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat 1200
cgatttcatg attttttttt gtttcgttgc atagggtttg gtttgccctt ttcctttatt 1260
tcaatatatg ccgtgcactt gtttgtcggg tcatcttttc atgtttttt tttggcttgg 1320
ttgtgatgat gtggtctggt tgggcggtcg ttctagatcg gagtagaatt ctgtttcaaa 1380
ctacctggtg gatttattaa ttttggatct gtatgtgtgt gccatacata ttcatagtta 1440
cgaattgaag atgatggatg gaaatatcga tctaggatag gtatacatgt tgatgcgggt 1500
tttactgatg catatacaga gatgcttttt gttcgcttgg ttgtgatgat gtggtgtggt 1560
tgggcggtcg ttcattcgtt ctagatcgga gtagaatact gtttcaaact aactggtgga 1620

[0012]

tttattaatt ttggatctgt atgtgtgtgc catacatctt catagttacg agtttaagat 1680

gatggatgga agtatcgatc taggataggt atacatgttg atgtgggttt tactgatgca 1740

tatacatgat ggcatatgca gcatctattc atatgctcta accttgagta cctatctatt 1800

ataataaaca agtatgtttt ataattattt tgatcttgat atacttggat gatggcatat 1860

gcagcagcta tatgtggatt tttttagccc tgccttcata cgctatttat ttgcttggta 1920

ctgtttcttt tgtcgatgct caccctgttg tttggtgtta cttctgcaga ccatggtgaa 1980

caaacacttc ttgtcccttt cggtcctcat cgtcctcctt ggcctctcct ccaacttgac 2040

agccgggggа tccaccatga aatatacaag ttatatcttg gcttttcagc tctgcatcgt 2100

tttgggttct cttggctgtt actgccagga cccatatgta aaagaagcag aaaaccttaa 2160

gaaatatttt aatgcaggtc attcagatgt agcggataat ggaactcttt tcttaggcat 2220

tttgaagaat tggaaagagg agagtgacag aaaaataatg cagagccaaa ttgtctcctt 2280

ttacttcaaa ctttttaaaa actttaaaga tgaccagagc atccaaaaga gtgtggagac 2340

catcaaggaa gacatgaatg tcaagttttt caatagcaac aaaaagaaac gagatgactt 2400

cgaaaagctg actaattatt cggtaactga cttgaatgtc caacgcaaag caatacatga 2460

actcatccaa gtgatggctg aactgtcgcc agcagctaaa acagggaagc gaaaaaggag 2520

tcagatgctg tttcgaggtc gaagagcatc ccagtaatgg ttgtcctgcc tgcaatattt 2580

gaattttaaa tctaaatcta tttattaagg atccggatcc gcccctctcc ctcccccccc 2640

cctaacgtta ctggccgaag ccgcttggaa taaggccggt gtgcgtttgt ctatatgtta 2700

ttttccacca tattgccgtc ttttggcaat gtgagggccc ggaaacctgg ccctgtcttc 2760

ttgacgagca ttcctagggg tctttccccт ctcgccaaag gaatgcaagg tctgttgaat 2820

gtcgtgaagg aagcagttcc tctggaagct tcttgaagac aaacaacgtc tgtagcgacc 2880

ctttgcaggc agcggaaccc cccacctggc gacaggtgcc tctgcggcca aaagccacgt 2940

gtataagata cacctgcaaa ggcggcacaa ccccagtgcc acgttgtgag ttggatagtt 3000

gtggaaagag tcaaatggct ctcctcaagc gtattcaaca aggggctgaa ggatgcccag 3060

aaggtacccc attgtatggg atctgatctg gggcctcggt gcacatgctt tacatgtgtt 3120

tagtcgaggt taaaaaaacg tctaggcccc ccgaaccacg gggacgtggt tttcctttga 3180

aaaacacgat gataatatgg ccacctcgag accatgattg aacaagatgg attgcacgca 3240

ggttctccgg ccgcttgggt ggagaggcta ttcggctatg actgggcaca acagacaatc 3300

ggctgctctg atgccgccgt gttccggctg tcagcgcagg ggcgcccggt tcttttgtc 3360

aagaccgacc tgtccggtgc cctgaatgaa ctgcaggacg aggcagcgcg gctatcgtgg 3420

ctggccacga cgggcgttcc ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg 3480

gactggctgc tattgggcga agtgccgggg caggatctcc tgtcatctca ccttgctcct 3540

[0013]

gccgagaaag tatccatcat ggctgatgca atgcggcggc tgcatacgct tgatccggct 3600

acctgcccat tcgaccacca agcgaaacat cgcatcgagc gagcacgtac tcggatggaa 3660

gccggtcttg tcgatcagga tgatctggac gaagagcatc aggggctcgc gccagccgaa 3720

ctgttcgcca ggctcaaggc gcgcatgccc gacggcgatg atctcgtcgt gacccatggc 3780

gatgcctgct tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt 3840

ggccggctgg gtgtggcgga ccgctatcag gacatagcgt tggctacccg tgatattgct 3900

gaagagcttg cggcgaatg ggctgaccgc ttcctcgtgc tttacggtat cgccgctccc 3960

gattcgcagc gcatcgcctt ctatcgcctt cttgacgagt tcttctgagc gggactctgg 4020

ggttcgaaat gaccgaccaa gcgacgccca acctgccatc acgagatttc gattccaccg 4080

ccgccttcta tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc 4140

tccagcgcgg ggatctcatg ctggagttct tcgcccacgg gatctctgcg gaacaggcgg 4200

tcgaaggtgc cgatatcatt acgacagcaa cggccgacaa gcacaacgcc acgatcctga 4260

gcgacaatat gatcgggccc ggcgtccaca tcaacggcgt cggcggcgac tgcccaggca 4320

agaccgagat gcaccgcgat atcttgctgc gttcggatat tttcgtggag ttcccgccac 4380

agacccggat gatccccgat cgttcaaaca tttggcaata aagtttctta agattgaatc 4440

ctgttgccgg tcttgcgatg attatcatat aatttctgtt gaattacgtt aagcatgtaa 4500

taattaacat gtaatgcatg acgttattta tgagatgggt ttttatgatt agagtcccgc 4560

aattatacat ttaatacgcg atagaaaaca aaatatagcg cgcaaactag gataaattat 4620

cgcgcgcggt gtcatctatg ttactagatc gggcctcctg tcaatgctgg cggcggctct 4680

ggtctaga 4688

<210> 38
<211> 4178
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 38
aagctttgat cccctaatga gcattgcatg tctaagttat aaaaaattac cacatatttt 60

tttgtcacac ttgtttgaag tgcagtttat ctatctttat acatatattt aaacttcact 120

ctacgaataa tataatctat agtactacaa taatatcagt gttttagaga atcatataaa 180

tgaacagtta gacatggtct aaaggacaat tgagtacttt gacaacagga ctctacagtt 240

ttatctcttt agtgtgcatg tgttctcctt ttttttttgca aatagcttca cctatataat 300

acttcatcca ttttattagt acatccattt agggtttagg gttaatggtt tttatagact 360

aatttttta gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact 420

ctattttagt ttttttattt aataatttag atataaaata gaataaaata aagtgactaa 480

[0014]

aaattaaaca aatacccttt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg 540

agtagataat gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa 600

ccagcagcgt cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct 660

gggcccctct cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa 720

ttgcgtggcg gagcggcaga cgtgagccgg cacggcaggc ggcctcctcc tcctctcacg 780

gcaccggcag ctacgggggа ttcctttccc accgctcctt cgctttccct tcctcgcccg 840

ccgtaataaa tagacacccc ctccacaccc tctttcccca acctcgtgtt gttcggagcg 900

cacacacaca caaccagatc tcccccaaat ccacccgtcg gcacctccgc ttcaaggtac 960

gccgctcgtc ctcccccccc ccctctctac cttctctaga tcggcgttcc ggtccatggt 1020

tagggcccgg tagttctact tctgttcatg tttgtgttag atccgtgttt gtgatagatc 1080

cgtgctgcta gcgttcgtac acggatgcga cctgtacgtc agacacgttc tgattgctaa 1140

cttgccagtg tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat 1200

cgatttcatg attttttttt gtttcgttgc atagggtttg gtttgccctt ttcctttatt 1260

tcaatatatg ccgtgcactt gtttgtcggg tcatcttttc atgtttttt tttggcttgg 1320

ttgtgatgat gtggtctggt tgggcggtcg ttctagatcg gagtagaatt ctgtttcaaa 1380

ctacctggtg gatttattaa ttttggatct gtatgtgtgt gccatacata ttcatagtta 1440

cgaattgaag atgatggatg gaaatatcga tctaggatag gtatacatgt tgatgcgggt 1500

tttactgatg catatacaga gatgcttttt gttcgcttgg ttgtgatgat gtggtgtggt 1560

tgggcggtcg ttcattcgtt ctagatcgga gtagaatact gtttcaaact aactggtgga 1620

tttattaatt ttggatctgt atgtgtgtgc catacatctt catagttacg agtttaagat 1680

gatggatgga agtatcgatc taggataggt atacatgttg atgtgggttt tactgatgca 1740

tatacatgat ggcatatgca gcatctattc atatgctcta accttgagta cctatctatt 1800

ataataaaca agtatgtttt ataattattt tgatcttgat atacttggat gatggcatat 1860

gcagcagcta tatgtggatt tttttagccc tgccttcata cgctatttat ttgcttggta 1920

ctgtttcttt tgtcgatgct caccctgttg tttggtgtta cttctgcagg atccaccatg 1980

gtgagcaagg gcgaggagct gttcaccggg gtggtgccca tcctggtcga gctggacggc 2040

gacgtaaacg gccacaagtt cagcgtgtcc ggcgagggcg agggcgatgc cacctacggc 2100

aagctgaccc tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gcccaccctc 2160

gtgaccaccc tgacctacgg cgtgcagtgc ttcagccgct accccgacca catgaagcag 2220

cacgacttct tcaagtccgc catgcccgaa ggctacgtcc aggagcgcac catcttcttc 2280

aaggacgacg gcaactacaa gacccgcgcc gaggtgaagt tcgagggcga caccctggtg 2340

aaccgcatcg agctgaaggg catcgacttc aaggaggacg gcaacatcct ggggcacaag 2400

[0015]

ctggagtaca actacaacag ccacaacgtc tatatcatgg ccgacaagca gaagaacggc 2460

atcaaggtga acttcaagat ccgccacaac atcgaggacg gcagcgtgca gctcgccgac 2520

cactaccagc agaacacccc catcggcgac ggccccgtgc tgctgcccga caaccactac 2580

ctgagcaccc agtccgccct gagcaaagac cccaacgaga agcgcgatca catggtcctg 2640

ctggagttcg tgaccgccgc cgggatcact ctcggcatgg acgagctgta caagctcgag 2700

accatgattg aacaagatgg attgcacgca ggttctccgg ccgcttgggt ggagaggcta 2760

ttcggctatg actgggcaca acagacaatc ggctgctctg atgccgccgt gttccggctg 2820

tcagcgcagg ggcgcccggt tcttttgtc aagaccgacc tgtccggtgc cctgaatgaa 2880

ctgcaggacg aggcagcgcg gctatcgtgg ctggccacga cgggcgttcc ttgcgcagct 2940

gtgctcgacg ttgtcactga agcgggaagg gactggctgc tattgggcga agtgccgggg 3000

caggatctcc tgtcatctca ccttgctcct gccgagaaag tatccatcat ggctgatgca 3060

atgcggcggc tgcatacgct tgatccggct acctgcccat tcgaccacca agcgaaacat 3120

cgcatcgagc gagcacgtac tcggatggaa gccggtcttg tcgatcagga tgatctggac 3180

gaagagcatc aggggctcgc gccagccgaa ctgttcgcca ggctcaaggc gcgcatgccc 3240

gacggcgatg atctcgtcgt gacccatggc gatgcctgct tgccgaatat catggtggaa 3300

aatggccgct tttctggatt catcgactgt ggccggctgg gtgtggcgga ccgctatcag 3360

gacatagcgt tggctacccg tgatattgct gaagagcttg gcggcgaatg ggctgaccgc 3420

ttcctcgtgc tttacggtat cgccgctccc gattcgcagc gcatcgcctt ctatcgcctt 3480

cttgacgagt tcttctgagc gggactctgg ggttcgaaat gaccgaccaa gcgacgccca 3540

acctgccatc acgagatttc gattccaccg ccgccttcta tgaaaggttg ggcttcggaa 3600

tcgttttccg ggacgccggc tggatgatcc tccagcgcgg ggatctcatg ctggagttct 3660

tcgcccacgg gatctctgcg gaacaggcgg tcgaaggtgc cgatatcatt acgacagcaa 3720

cggccgacaa gcacaacgcc acgatcctga gcgacaatat gatcgggccc ggcgtccaca 3780

tcaacggcgt cggcggcgac tgcccaggca agaccgagat gcaccgcgat atcttgctgc 3840

gttcggatat tttcgtggag ttcccgccac agacccggat gatccccgat cgttcaaaca 3900

tttggcaata aagtttctta agattgaatc ctgttgccgg tcttgcgatg attatcatat 3960

aatttctgtt gaattacgtt aagcatgtaa taattaacat gtaatgcatg acgttattta 4020

tgagatgggt ttttatgatt agagtcccgc aattatacat ttaatacgcg atagaaaaca 4080

aaatatagcg cgcaaactag gataaattat cgcgcgcggt gtcatctatg ttactagatc 4140

gggcctcctg tcaatgctgg cggcggctct ggtctaga 4178

<210> 39
<211> 4663
<212> DNA
<213> 人工序列

[0016]

<220>
<223> 合成寡核苷酸

<400> 39
aagctttgat cccctaatga gcattgcatg tctaagttat aaaaaattac cacatatttt 60

tttgtcacac ttgtttgaag tgcagtttat ctatctttat acatatattt aaacttcact 120

ctacgaataa tataatctat agtactacaa taatatcagt gttttagaga atcatataaa 180

tgaacagtta gacatggtct aaaggacaat tgagtacttt gacaacagga ctctacagtt 240

ttatctcttt agtgtgcatg tgttctcctt ttttttgca aatagcttca cctatataat 300

acttcatcca ttttattagt acatccattt agggtttagg gttaatggtt tttatagact 360

aatttttta gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact 420

ctattttagt ttttttattt aataatttag atataaaata gaataaaata aagtgactaa 480

aaattaaaca aatacccttt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg 540

agtagataat gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa 600

ccagcagcgt cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct 660

gggcccctct cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa 720

ttgcgtggcg gagcggcaga cgtgagccgg cacggcaggc ggcctcctcc tcctctcacg 780

gcaccggcag ctacgggggа ttcctttccc accgctcctt cgctttcccc tcctcgcccg 840

ccgtaataaa tagacacccc ctccacaccc tctttccccа acctcgtgtt gttcggagcg 900

cacacacaca caaccagatc tcccccaaat ccacccgtcg gcacctccgc ttcaaggtac 960

gccgctcgtc ctcccccccc ccctctctac cttctctaga tcggcgttcc ggtccatggt 1020

tagggcccgg tagttctact tctgttcatg tttgtgttag atccgtgttt gtgatagatc 1080

cgtgctgcta gcgttcgtac acggatgcga cctgtacgtc agacacgttc tgattgctaa 1140

cttgccagtg tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat 1200

cgatttcatg attttttttt gtttcgttgc atagggtttg gtttgccctt ttcctttatt 1260

tcaatatatg ccgtgcactt gtttgtcggg tcatcttttc atgtttttt tttggcttgg 1320

ttgtgatgat gtggtctggt tgggcggtcg ttctagatcg gagtagaatt ctgtttcaaa 1380

ctacctggtg gatttattaa ttttggatct gtatgtgtgt gccatacata ttcatagtta 1440

cgaattgaag atgatggatg gaaatatcga tctaggatag gtatacatgt tgatgcgggt 1500

tttactgatg catatacaga gatgcttttt gttcgcttgg ttgtgatgat gtggtgtggt 1560

tgggcggtcg ttcattcgtt ctagatcgga gtagaatact gtttcaaact aactggtgga 1620

tttattaatt ttggatctgt atgtgtgtgc catacatctt catagttacg agtttaagat 1680

gatggatgga agtatcgatc taggataggt atacatgttg atgtgggttt tactgatgca 1740

tatacatgat ggcatatgca gcatctattc atatgctcta accttgagta cctatctatt 1800

[0017]

ataataaaca agtatgtttt ataattattt tgatcttgat atacttggat gatggcatat 1860

gcagcagcta tatgtggatt tttttagccc tgccttcata cgctatttat ttgcttggta 1920

ctgtttcttt tgtcgatgct caccctgttg tttggtgtta cttctgcagg atccaacaat 1980

ggccttgacc tttgctttac tggtggccct cctggtgctc agctgcaagt caagctgctc 2040

tgtgggctgt gatctgcctc aaacccacag cctgggtagc aggaggacct tgatgctcct 2100

ggcacagatg aggagaatct ctcttttctc ctgcttgaag gacagacatg actttggatt 2160

tccccaggag gagtttggca accagttcca aaaggctgaa accatccctg tcctccatga 2220

gatgatccag cagatcttca atctcttcag cacaaaggac tcatctgctg cttgggatga 2280

gaccctccta gacaaattct acactgaact ctaccagcag ctgaatgacc tggaagcctg 2340

tgtgatacag ggggtggggg tgacagagac tcccctgatg aaggaggact ccattctggc 2400

tgtgaggaaa tacttccaaa gaatcactct ctatctgaaa gagaagaaat acagcccttg 2460

tgcctgggag gttgtcagag cagaaatcat gagatctttt tctttgtcaa caaacttgca 2520

agaaagttta agaagtaagg aatgagagct caagcttcga attctgcagt cgacggtacc 2580

gcgggcccgg gatccgcccc tctccctccc ccccccctaa cgttactggc cgaagccgct 2640

tggaataagg ccggtgtgcg tttgtctata tgttattttc caccatattg ccgtcttttg 2700

gcaatgtgag ggcccggaaa cctggccctg tcttcttgac gagcattcct aggggtcttt 2760

cccctctcgc caaaggaatg caaggtctgt tgaatgtcgt gaaggaagca gttcctctgg 2820

aagcttcttg aagacaaaca acgtctgtag cgacccttta caggcagcgg aaccccccac 2880

ctggcgacag gtgcctctgc ggccaaaagc cacgtgtata agatacacct gcaaaggcgg 2940

cacaacccca gtgccacgtt gtgagttgga tagttgtgga aagagtcaaa tggctctcct 3000

caagcgtatt caacaagggg ctgaaggatg cccagaaggt accccattgt atgggatctg 3060

atctggggcc tcggtgcaca tgctttacat gtgtttagtc gaggttaaaa aaacgtctag 3120

gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgatgataa tatggccacc 3180

tcgagaccat gattgaacaa gatggattgc acgcaggttc tccggccgct tgggtggaga 3240

ggctattcgg ctatgactgg gcacaacaga caatcggctg ctctgatgcc gccgtgttcc 3300

ggctgtcagc gcaggggcgc ccggttcttt ttgtcaagac cgacctgtcc ggtgccctga 3360

atgaactgca ggacgaggca gcgcggctat cgtggctggc cacgacgggc gttccttgcg 3420

cagctgtgct cgacgttgtc actgaagcgg gaagggactg gctgctattg ggcgaagtgc 3480

cggggcagga tctcctgtca tctcaccttg ctcctgccga gaaagtatcc atcatggctg 3540

atgcaatgcg gcggctgcat acgcttgatc cggctacctg cccattcgac caccaagcga 3600

aacatcgcat cgagcgagca cgtactcgga tggaagccgg tcttgtcgat caggatgatc 3660

tggacgaaga gcatcagggg ctcgcgccag ccgaactgtt cgccaggctc aaggcgcgca 3720

tgcccgacgg cgatgatctc gtcgtgaccc atggcgatgc ctgcttgccg aatatcatgg 3780

[0018]

tggaaaatgg ccgcttttct ggattcatcg actgtggccg gctgggtgtg gcggaccgct 3840

atcaggacat agcgttggct acccgtgata ttgctgaaga gcttggcggc gaatgggctg 3900

accgcttcct cgtgctttac ggtatcgccg ctcccgattc gcagcgcatc gccttctatc 3960

gccttcttga cgagttcttc tgagcgggac tctggggttc gaaatgaccg accaagcgac 4020

gcccaacctg ccatcacgag atttcgattc caccgccgcc ttctatgaaa ggttgggctt 4080

cggaatcgtt ttccgggacg ccggctggat gatcctccag cgcggggatc tcatgctgga 4140

gttcttcgcc cacgggatct ctgcggaaca ggcggtcgaa ggtgccgata tcattacgac 4200

agcaacggcc gacaagcaca acgccacgat cctgagcgac aatatgatcg ggcccggcgt 4260

ccacatcaac ggcgtcggcg gcgactgccc aggcaagacc gagatgcacc gcgatatctt 4320

gctgcgttcg gatattttcg tggagttccc gccacagacc cggatgatcc ccgatcgttc 4380

aaacatttgg caataaagtt tcttaagatt gaatcctgtt gccggtcttg cgatgattat 4440

catataattt ctgttgaatt acgttaagca tgtaataatt aacatgtaat gcatgacgtt 4500

atttatgaga tgggttttta tgattagagt cccgcaatta tacatttaat acgcgataga 4560

aaacaaaata tagcgcgcaa actaggataa attatcgcgc gcggtgtcat ctatgttact 4620

agatcgggcc tcctgtcaat gctggcggcg gctctggtct aga 4663

<210> 40
<211> 4163
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 40
aagctttgat ccctaatga gcattgcatg tctaagttat aaaaaattac cacatatttt 60

tttgtcacac ttgtttgaag tgcagtttat ctatctttat acatatattt aaacttcact 120

ctacgaataa tataatctat agtactacaa taatatcagt gttttagaga atcatataaa 180

tgaacagtta gacatggtct aaaggacaat tgagtacttt gacaacagga ctctacagtt 240

ttatctcttt agtgtgcatg tgttctcctt ttttttgca aatagcttca cctatataat 300

acttcatcca ttttattagt acatccattt agggtttagg gttaatggtt tttatagact 360

aatttttta gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact 420

ctattttagt ttttttattt aataatttag atataaaata gaataaaata aagtgactaa 480

aaattaaaca aataccctt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg 540

agtagataat gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa 600

ccagcagcgt cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct 660

gggcccctct cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa 720

[0019]

ttgcgtggcg gagcggcaga cgtgagccgg cacggcaggc ggcctcctcc tcctctcacg 780

gcaccggcag ctacgggggа ttcctttccc accgctcctt cgctttcccт tcctcgcccg 840

ccgtaataaa tagacacccc ctccacaccc tctttcccca acctcgtgtt gttcggagcg 900

cacacacaca caaccagatc tcccccaaat ccacccgtcg gcacctccgc ttcaaggtac 960

gccgctcgtc ctcccccccc ccctctctac cttctctaga tcggcgttcc ggtccatggt 1020

tagggcccgg tagttctact tctgttcatg tttgtgttag atccgtgttt gtgatagatc 1080

cgtgctgcta gcgttcgtac acggatgcga cctgtacgtc agacacgttc tgattgctaa 1140

cttgccagtg tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat 1200

cgatttcatg attttttttt gtttcgttgc atagggtttg gtttgccctt ttcctttatt 1260

tcaatatatg ccgtgcactt gtttgtcggg tcatctttтc atgtttтttt tttggcttgg 1320

ttgtgatgat gtggtctggt tgggcggtcg ttctagatcg gagtagaatt ctgtttcaaa 1380

ctacctggtg gatttattaa ttttggatct gtatgtgtgt gccatacata ttcatagtta 1440

cgaattgaag atgatggatg gaaatatcga tctaggatag gtatacatgt tgatgcgggt 1500

tttactgatg catatacaga gatgcttttt gttcgcttgg ttgtgatgat gtggtgtggt 1560

tgggcggtcg ttcattcgtt ctagatcgga gtagaatact gtttcaaact aactggtgga 1620

tttattaatt ttggatctgt atgtgtgtgc catacatctt catagttacg agtttaagat 1680

gatggatgga agtatcgatc taggataggt atacatgttg atgtgggttt tactgatgca 1740

tatacatgat ggcatatgca gcatctattc atatgctcta accttgagta cctatctatt 1800

ataataaaca agtatgtttt ataattattt tgatcttgat atacttggat gatggcatat 1860

gcagcagcta tatgtggatt tttttagccc tgccttcata cgctatttat ttgcttggta 1920

ctgtttcttt tgtcgatgct caccctgttg tttggtgtta cttctgcaga ccatggtgaa 1980

caaacacttc ttgtcccttt cggtcctcat cgtcctcctt ggcctctcct ccaacttgac 2040

agccgggggа tccaccatga ggttcatgac tctcctcttc ctgacagctc tggcaggagc 2100

cctggtctgt gcctatgatc cagaggccgc ctctgcccca ggatcgggga acccttgcca 2160

tgaagcatca gcagctcaaa aggaaaatgc aggtgaagac ccagggttag ccagacaggc 2220

accaaagcca aggaagcaga gatccagcct tctggaaaaa ggcctagacg gagcaaaaaa 2280

agctgtgggg ggactcggaa aactaggaaa agatgcagtc gaagatctag aaagcgtggg 2340

taaaggagcc gtccatgacg ttaaagacgt ccttgactca gtactatagg gatccgcccc 2400

tctccctccc ccccccctaa cgttactggc cgaagccgct tggaataagg ccggtgtgcg 2460

tttgtctata tgttattttc caccatattg ccgtcttttg gcaatgtgag ggcccggaaa 2520

cctggccctg tcttcttgac gagcattcct aggggtcttt cccctctcgc caaaggaatg 2580

caaggtctgt tgaatgtcgt gaaggaagca gttcctctgg aagcttcttg aagacaaaca 2640

acgtctgtag cgacccttтg caggcagcgg aacccccccac ctggcgacag gtgcctctgc 2700

[0020]

ggccaaaagc cacgtgtata agatacacct gcaaaggcgg cacaacccca gtgccacgtt 2760

gtgagttgga tagttgtgga aagagtcaaa tggctctcct caagcgtatt caacaagggg 2820

ctgaaggatg cccagaaggt accccattgt atgggatctg atctggggcc tcggtgcaca 2880

tgctttacat gtgtttagtc gaggttaaaa aaacgtctag gccccccgaa ccacggggac 2940

gtggttttcc tttgaaaaac acgatgataa tatggccacc tcgagaccat gattgaacaa 3000

gatggattgc acgcaggttc tccggccgct tgggtggaga ggctattcgg ctatgactgg 3060

gcacaacaga caatcggctg ctctgatgcc gccgtgttcc ggctgtcagc gcaggggcgc 3120

ccggttcttt ttgtcaagac cgacctgtcc ggtgccctga atgaactgca ggacgaggca 3180

gcgcggctat cgtggctggc cacgacgggc gttccttgcg cagctgtgct cgacgttgtc 3240

actgaagcgg gaagggactg gctgctattg ggcgaagtgc cggggcagga tctcctgtca 3300

tctcaccttg ctcctgccga gaaagtatcc atcatggctg atgcaatgcg gcggctgcat 3360

acgcttgatc cggctacctg cccattcgac caccaagcga aacatcgcat cgagcgagca 3420

cgtactcgga tggaagccgg tcttgtcgat caggatgatc tggacgaaga gcatcagggg 3480

ctcgcgccag ccgaactgtt cgccaggctc aaggcgcgca tgcccgacgg cgatgatctc 3540

gtcgtgaccc atggcgatgc ctgcttgccg aatatcatgg tggaaaatgg ccgcttttct 3600

ggattcatcg actgtggccg gctgggtgtg gcggaccgct atcaggacat agcgttggct 3660

acccgtgata ttgctgaaga gcttggcggc gaatgggctg accgcttcct cgtgctttac 3720

ggtatcgccg ctcccgattc gcagcgcatc gccttctatc gccttcttga cgagttcttc 3780

tgagcgggac tctggggttc gaaatgaccg accaagcgac gcccaacctg ccatcacgag 3840

atttcgattc caccgccgcc ttctatgaaa ggttgggctt cggaatcgtt ttccgggacg 3900

ccggctggat gatcctccag cgcggggatc tcatgctgga gttcttcgcc cacgggatct 3960

ctgcggaaca ggcggtcgaa ggtgccgata tcattacgac agcaacggcc gacaagcaca 4020

acgccacgat cctgagcgac aatatgatcg ggcccggcgt ccacatcaac ggcgtcggcg 4080

gcgactgccc aggcaagacc gagatgcacc gcgatatctt gctgcgttcg gatatttcg 4140

tggagttccc gccacagacc cggatgatcc ccgatcgttc aaacatttgg caataaagtt 4200

tcttaagatt gaatcctgtt gccggtcttg cgatgattat catataattt ctgttgaatt 4260

acgttaagca tgtaataatt aacatgtaat gcatgacgtt atttatgaga tgggttttta 4320

tgattagagt cccgcaatta tacatttaat acgcgataga aaacaaaata tagcgcgcaa 4380

actaggataa attatcgcgc gcggtgtcat ctatgttact agatcgggcc tcctgtcaat 4440

gctggcggcg gctctggtct aga 4463

<210> 41
<211> 4784
<212> DNA

[0021]

<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 41
aagctttgat cccctaatga gcattgcatg tctaagttat aaaaaattac cacatatttt 60

tttgtcacac ttgtttgaag tgcagtttat ctatctttat acatatattt aaacttcact 120

ctacgaataa tataatctat agtactacaa taatatcagt gttttagaga atcatataaa 180

tgaacagtta gacatggtct aaaggacaat tgagtacttt gacaacagga ctctacagtt 240

ttatctcttt agtgtgcatg tgttctcctt ttttttgca aatagcttca cctatataat 300

acttcatcca ttttattagt acatccattt agggtttagg gttaatggtt tttatagact 360

aatttttta gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact 420

ctattttagt ttttttattt aataatttag atataaaata gaataaaata aagtgactaa 480

aaattaaaca aatacccttt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg 540

agtagataat gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa 600

ccagcagcgt cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct 660

gggcccctct cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa 720

ttgcgtggcg gagcggcaga cgtgagccgg cacggcaggc ggcctcctcc tcctctcacg 780

gcaccggcag ctacggggga ttcctttccc accgctcctt cgctttccct tcctcgcccg 840

ccgtaataaa tagacacccc ctccacaccc tctttcccca acctcgtgtt gttcggagcg 900

cacacacaca caaccagatc tcccccaaat ccacccgtcg gcacctccgc ttcaaggtac 960

gccgctcgtc ctcccccccc ccctctctac cttctctaga tcggcgttcc ggtccatggt 1020

tagggcccgg tagttctact tctgttcatg tttgtgttag atccgtgttt gtgatagatc 1080

cgtgctgcta gcgttcgtac acggatgcga cctgtacgtc agacacgttc tgattgctaa 1140

cttgccagtg tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat 1200

cgatttcatg attttttttt gtttcgttgc atagggtttg gtttgccctt ttcctttatt 1260

tcaatatatg ccgtgcactt gtttgtcggg tcatcttttc atgttttttt tttggcttgg 1320

ttgtgatgat gtggtctggt tgggcggtcg ttctagatcg gagtagaatt ctgtttcaaa 1380

ctacctggtg gatttattaa ttttggatct gtatgtgtgt gccatacata ttcatagtta 1440

cgaattgaag atgatggatg gaaatatcga tctaggatag gtatacatgt tgatgcgggt 1500

tttactgatg catatacaga gatgcttttt gttcgcttgg ttgtgatgat gtggtgtggt 1560

tgggcggtcg ttcattcgtt ctagatcgga gtagaatact gtttcaaact aactggtgga 1620

tttattaatt ttggatctgt atgtgtgtgc catacatctt catagttacg agtttaagat 1680

gatggatgga agtatcgatc taggataggt atacatgttg atgtgggttt tactgatgca 1740

tatacatgat ggcatatgca gcatctattc atatgctcta accttgagta cctatctatt 1800

[0022]

ataataaaca agtatgtttt ataattattt tgatcttgat atacttggat gatggcatat 1860

gcagcagcta tatgtggatt ttttttagccc tgccttcata cgctatttat ttgcttggta 1920

ctgtttcttt tgtcgatgct caccctgttg tttggtgtta cttctgcaga ccatggtgaa 1980

caaacacttc ttgtcccttt cggtcctcat cgtcctcctt ggcctctcct ccaacttgac 2040

agccgggggа tccaccatgg ctacaggctc ccggacgtcc ctgctcctgg cttttggcct 2100

gctctgcctg ccctggcttc aagagggcag tgccttccca accattccct tatccaggct 2160

ttttgacaac gctatgctcc gcgcccatcg tctgcaccag ctggcctttg acacctacca 2220

ggagtttgaa gaagcctata tcccaaagga acagaagtat tcattcctgc agaacccccа 2280

gacctccctc tgtttctcag agtctattcc gacaccctcc aacagggagg aaacacaaca 2340

gaaatccaac ctagagctgc tccgcatctc cctgctgctc atccagtcgt ggctggagcc 2400

cgtgcagttc ctcaggagtg tcttcgccaa cagcctggtg tacggcgcct ctgacagcaa 2460

cgtctatgac ctcctaaagg acctagagga aggcatccaa acgctgatgg ggaggctgga 2520

agatggcagc ccccggactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa 2580

ctcacacaac gatgacgcac tactcaagaa ctacgggctg ctctactgct tcaggaagga 2640

catggacaag gtcgagacat tcctgcgcat cgtgcagtgc cgctctgtgg agggcagctg 2700

tggcttctag ggatccgccc ctctccctcc ccccccccta acgttactgg ccgaagccgc 2760

ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt gccgtctttt 2820

ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc taggggtctt 2880

tcccctctcg ccaaaggaat gcaaggtctg ttgaatgtcg tgaaggaagc agttcctctg 2940

gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg gaaccccccа 3000

cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc tgcaaaggcg 3060

gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa atggctctcc 3120

tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg tatgggatct 3180

gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa aaaacgtcta 3240

ggcccccccga accacgggga cgtggttttc ctttgaaaaa cacgatgata atatggccac 3300

ctcgagacca tgattgaaca agatggattg cacgcaggtt ctccggccgc ttgggtggag 3360

aggctattcg gctatgactg ggcacaacag acaatcggct gctctgatgc cgccgtgttc 3420

cggctgtcag cgcaggggcg cccggttctt tttgtcaaga ccgacctgtc cggtgccctg 3480

aatgaactgc aggacgaggc agcgcggcta tcgtggctgg ccacgacggg cgttccttgc 3540

gcagctgtgc tcgacgttgt cactgaagcg ggaagggact ggctgctatt gggcgaagtg 3600

ccggggcagg atctcctgtc atctcacctt gctcctgccg agaaagtatc catcatggct 3660

gatgcaatgc ggcggctgca tacgcttgat ccggctacct gcccattcga ccaccaagcg 3720

[0023]

aaacatcgca tcgagcgagc acgtactcgg atggaagccg gtcttgtcga tcaggatgat 3780

ctggacgaag agcatcaggg gctcgcgcca gccgaactgt tcgccaggct caaggcgcgc 3840

atgcccgacg gcgatgatct cgtcgtgacc catggcgatg cctgcttgcc gaatatcatg 3900

gtggaaaatg gccgcttttc tggattcatc gactgtggcc ggctgggtgt ggcggaccgc 3960

tatcaggaca tagcgttggc tacccgtgat attgctgaag agcttggcgg cgaatgggct 4020

gaccgcttcc tcgtgcttta cggtatcgcc gctcccgatt cgcagcgcat cgccttctat 4080

cgccttcttg acgagttctt ctgagcggga ctctggggtt cgaaatgacc gaccaagcga 4140

cgcccaacct gccatcacga gatttcgatt ccaccgccgc cttctatgaa aggttgggct 4200

tcggaatcgt tttccgggac gccggctgga tgatcctcca gcgcggggat ctcatgctgg 4260

agttcttcgc ccacgggatc tctgcggaac aggcggtcga aggtgccgat atcattacga 4320

cagcaacggc cgacaagcac aacgccacga tcctgagcga caatatgatc gggccccggcg 4380

tccacatcaa cggcgtcggc ggcgactgcc caggcaagac cgagatgcac cgcgatatct 4440

tgctgcgttc ggatattttc gtggagttcc cgccacagac ccggatgatc cccgatcgtt 4500

caaacatttg gcaataaagt ttcttaagat tgaatcctgt tgccggtctt gcgatgatta 4560

tcatataatt tctgttgaat tacgttaagc atgtaataat taacatgtaa tgcatgacgt 4620

tatttatgag atgggttttt atgattagag tcccgcaatt atacatttaa tacgcgatag 4680

aaaacaaaat atagcgcgca aactaggata aattatcgcg cgcggtgtca tctatgttac 4740

tagatcgggc ctcctgtcaa tgctggcggc ggctctggtc taga 4784

<210> 42
<211> 5991
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 42
aagctttgat cccctaatga gcattgcatg tctaagttat aaaaaattac cacatatttt 60

tttgtcacac ttgtttgaag tgcagtttat ctatctttat acatatattt aaacttcact 120

ctacgaataa tataatctat agtactacaa taatatcagt gttttagaga atcatataaa 180

tgaacagtta gacatggtct aaaggacaat tgagtacttt gacaacagga ctctacagtt 240

ttatctcttt agtgtgcatg tgttctcctt ttttttgca aatagcttca cctatataat 300

acttcatcca ttttattagt acatccattt agggtttagg gttaatggtt tttatagact 360

aatttttta gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact 420

ctattttagt ttttttattt aataatttag atataaaata gaataaaata aagtgactaa 480

aaattaaaca aataccctttt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg 540

agtagataat gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa 600

[0024]

ccagcagcgt cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct 660

gggcccctct cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa 720

ttgcgtggcg gagcggcaga cgtgagccgg cacggcaggc ggcctcctcc tcctctcacg 780

gcaccggcag ctacgggggа ttcctttccc accgctcctt cgctttcccт tcctcgcccg 840

ccgtaataaa tagacacccc ctccacaccc tctttccccа acctcgtgtt gttcggagcg 900

cacacacaca caaccagatc tcccccaaat ccacccgtcg gcacctccgc ttcaaggtac 960

gccgctcgtc ctcccccccc ccctctctac cttctctaga tcggcgttcc ggtccatggt 1020

tagggcccgg tagttctact tctgttcatg tttgtgttag atccgtgttt gtgatagatc 1080

cgtgctgcta gcgttcgtac acggatgcga cctgtacgtc agacacgttc tgattgctaa 1140

cttgccagtg tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat 1200

cgatttcatg attttttttt gtttcgttgc atagggtttg gtttgcccтt ttcctttatt 1260

tcaatatatg ccgtgcactt gtttgtcggg tcatcttttc atgtttтttt tttggcttgg 1320

ttgtgatgat gtggtctggt tgggcggtcg ttctagatcg gagtagaatt ctgtttcaaa 1380

ctacctggtg gatttattaa ttttggatct gtatgtgtgt gccatacata ttcatagtta 1440

cgaattgaag atgatggatg gaaatatcga tctaggatag gtatacatgt tgatgcgggt 1500

tttactgatg catatacaga gatgcttttt gttcgcttgg ttgtgatgat gtggtgtggt 1560

tgggcggtcg ttcattcgtt ctagatcgga gtagaatact gtttcaaact aactggtgga 1620

tttattaatt ttggatctgt atgtgtgtgc catacatctt catagttacg agtttaagat 1680

gatggatgga agtatcgatc taggataggt atacatgttg atgtgggttt tactgatgca 1740

tatacatgat ggcatatgca gcatctattc atatgctcta accttgagta cctatctatt 1800

ataataaaca agtatgtttt ataattattt tgatcttgat atacttggat gatggcatat 1860

gcagcagcta tatgtggatt tttttagccc tgccttcata cgctatttat ttgcttggta 1920

ctgtttcttt tgtcgatgct caccctgttg tttggtgtta cttctgcaga ccatggtgaa 1980

caaacacttc ttgtcccttt cggtcctcat cgtcctcctt ggcctctcct ccaacttgac 2040

agccggggaa accatggcgc acgtccgagg cttgcagctg cctggctgcc tggccctggc 2100

tgccctgtgt agccttgtgc acagccagca tgtgttcctg gctcctcagc aagcacggtc 2160

gctgctccag cgggtccggc gagccaacac cttcttggag gaggtgcgca agggcaacct 2220

agagcgagag tgcgtggagg agacgtgcag ctacgaggag gccttcgagg ctctggagtc 2280

ctccacggct acggatgtgt tctgggccaa gtacacagct tgtgagacag cgaggacgcc 2340

tcgagataag cttgctgcat gtctggaagg taactgtgct gagggtctgg gtacgaacta 2400

ccgagggcat gtgaacatca cccggtcagg cattgagtgc cagctatgga ggagtcgcta 2460

cccacataag cctgaaatca actccactac ccatcctggg gccgacctac aggagaattt 2520

[0025]

ctgccgcaac cccgacagca gcaccacggg accctggtgc tacactacag accccaccgt 2580

gaggaggcag gaatgcagca tccctgtctg tggccaggat caagtcactg tagcgatgac 2640

tccacgctcc gaaggctcca gtgtgaatct gtcacctcca ttggagcagt gtgtccctga 2700

tcgggggcag cagtaccagg ggcgcctggc ggtgaccaca catgggctcc cctgcctggc 2760

ctgggccagc gcacaggcca aggccctgag caagcaccag gacttcaact cagctgtgca 2820

gctggtggag aacttctgcc gcaacccaga cggggatgag gagggcgtgt ggtgctatgt 2880

ggccgggaag cctggcgact ttgggtactg cgacctcaac tattgtgagg aggccgtgga 2940

ggaggagaca ggagatgggc tggatgagga ctcagacagg gccatcgaag ggcgtaccgc 3000

caccagtgag taccagactt tcttcaatcc gaggaccttt ggctcgggag aggcagactg 3060

tgggctgcga cctctgttcg agaagaagtc gctggaggac aaaaccgaaa gagagctcct 3120

ggaatcctac atcgacgggc gcattgtgga gggctcggat gcagagatcg gcatgtcacc 3180

ttggcaggtg atgcttttcc ggaagagtcc ccaggagctg ctgtgtgggg ccagcctcat 3240

cagtgaccgc tgggtcctca ccgccgccca ctgcctcctg tacccgccct gggacaagaa 3300

cttcaccgag aatgaccttc tggtgcgcat tggcaagcac tcccgcacaa ggtacgagcg 3360

aaacattgaa aagatatcca tgttggaaaa gatctacatc caccccaggt acaactggcg 3420

ggagaacctg gaccgggaca ttgccctgat gaagctgaag aagcctgttg ccttcagtga 3480

ctacattcac cctgtgtgtc tgcccgacag ggagacggca gccagcttgc tccaggctgg 3540

atacaagggg cgggtgacag gctggggcaa cctgaaggag acgtggacag ccaacgttgg 3600

taaggggcag cccagtgtcc tgcaggtggt gaacctgccc attgtggagc ggccggtctg 3660

caaggactcc acccggatcc gcatcactga caacatgttc tgtgctggtt acaagcctga 3720

tgaagggaaa cgaggggatg cctgtgaagg tgacagtggg ggaccctttg tcatgaagag 3780

ccccttaac aaccgctggt atcaaatggg catcgtctca tggggtgaag gctgtgaccg 3840

ggatgggaaa tatggcttct acacacatgt gttccgcctg aagaagtgga tacagaaggt 3900

cattgatcag tttggagagt agcgcccctc tccctccccc ccccctaacg ttactggccg 3960

aagccgcttg gaataaggcc ggtgtgcgtt tgtctatatg ttattttcca ccatattgcc 4020

gtcttttggc aatgtgaggg cccggaaacc tggccctgtc ttcttgacga gcattcctag 4080

gggtctttcc cctctcgcca aaggaatgca aggtctgttg aatgtcgtga aggaagcagt 4140

tcctctggaa gcttcttgaa gacaaacaac gtctgtagcg acccctttgca ggcagcggaa 4200

ccccccacct ggcgacaggt gcctctgcgg ccaaaagcca cgtgtataag atacacctgc 4260

aaaggcggca caaccccagt gccacgttgt gagttggata gttgtggaaa gagtcaaatg 4320

gctctcctca agcgtattca acaaggggct gaaggatgcc cagaaggtac cccattgtat 4380

gggatctgat ctggggcctc ggtgcacatg ctttacatgt gtttagtcga ggttaaaaaa 4440

acgtctaggc cccccgaacc acggggacgt ggttttcctt tgaaaaacac gatgataata 4500

[0026]

tggccacctc gagaccatga ttgaacaaga tggattgcac gcaggttctc cggccgcttg 4560

ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct ctgatgccgc 4620

cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg acctgtccgg 4680

tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca cgacgggcgt 4740

tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc tgctattggg 4800

cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga aagtatccat 4860

catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc cattcgacca 4920

ccaagcgaaa catcgcatcg agcgagcacg tactcggatg gaagccggtc ttgtcgatca 4980

ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg ccaggctcaa 5040

ggcgcgcatg cccgacggcg atgatctcgt cgtgacccat ggcgatgcct gcttgccgaa 5100

tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc tgggtgtggc 5160

ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc ttggcggcga 5220

atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc agcgcatcgc 5280

cttctatcgc cttcttgacg agttcttctg agcgggactc tggggttcga aatgaccgac 5340

caagcgacgc ccaacctgcc atcacgagat ttcgattcca ccgccgcctt ctatgaaagg 5400

ttgggcttcg gaatcgtttt ccgggacgcc ggctggatga tcctccagcg cggggatctc 5460

atgctggagt tcttcgccca cgggatctct gcggaacagg cggtcgaagg tgccgatatc 5520

attacgacag caacggccga caagcacaac gccacgatcc tgagcgacaa tatgatcggg 5580

cccggcgtcc acatcaacgg cgtcggcggc gactgcccag gcaagaccga gatgcaccgc 5640

gatatcttgc tgcgttcgga tattttcgtg gagttcccgc cacagacccg gatgatcccc 5700

gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc cggtcttgcg 5760

atgattatca tataatttct gttgaattac gttaagcatg taataattaa catgtaatgc 5820

atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata catttaatac 5880

gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcgc ggtgtcatct 5940

atgttactag atcgggcctc ctgtcaatgc tggcggcggc tctggtctag a 5991

<210> 43
<211> 4888
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 43
aagctttgat cccctaatga gcattgcatg tctaagttat aaaaaattac cacatatttt 60

tttgtcacac ttgtttgaag tgcagtttat ctatctttat acatatattt aaacttcact 120

[0027]

ctacgaataa tataatctat agtactacaa taatatcagt gttttagaga atcatataaa 180

tgaacagtta gacatggtct aaaggacaat tgagtacttt gacaacagga ctctacagtt 240

ttatctcttt agtgtgcatg tgttctcctt ttttttgca aatagcttca cctatataat 300

acttcatcca ttttattagt acatccattt agggtttagg gttaatggtt tttatagact 360

aatttttta gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact 420

ctattttagt ttttttattt aataatttag atataaaata gaataaaata aagtgactaa 480

aaattaaaca aatacccttt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg 540

agtagataat gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa 600

ccagcagcgt cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct 660

gggcccctct cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa 720

ttgcgtggcg gagcggcaga cgtgagccgg cacggcaggc ggcctcctcc tcctctcacg 780

gcaccggcag ctacggggga ttcctttccc accgctcctt cgctttcccc tcctcgcccg 840

ccgtaataaa tagacacccc ctccacaccc tctttccccа acctcgtgtt gttcggagcg 900

cacacacaca caaccagatc tcccccaaat ccacccgtcg gcacctccgc ttcaaggtac 960

gccgctcgtc ctcccccccc ccctctctac cttctctaga tcggcgttcc ggtccatggt 1020

tagggcccgg tagttctact tctgttcatg tttgtgttag atccgtgttt gtgatagatc 1080

cgtgctgcta gcgttcgtac acggatgcga cctgtacgtc agacacgttc tgattgctaa 1140

cttgccagtg tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat 1200

cgatttcatg attttttttt gtttcgttgc atagggtttg gtttgccctt ttcctttatt 1260

tcaatatatg ccgtgcactt gtttgtcggg tcatcttttc atgtttttt tttggcttgg 1320

ttgtgatgat gtggtctggt tgggcggtcg ttctagatcg gagtagaatt ctgtttcaaa 1380

ctacctggtg gatttattaa ttttggatct gtatgtgtgt gccatacata ttcatagtta 1440

cgaattgaag atgatggatg gaaatatcga tctaggatag gtatacatgt tgatgcgggt 1500

tttactgatg catatacaga gatgcttttt gttcgcttgg ttgtgatgat gtggtgtggt 1560

tgggcggtcg ttcattcgtt ctagatcgga gtagaatact gtttcaaact aactggtgga 1620

tttattaatt ttggatctgt atgtgtgtgc catacatctt catagttacg agtttaagat 1680

gatggatgga agtatcgatc taggataggt atacatgttg atgtgggttt tactgatgca 1740

tatacatgat ggcatatgca gcatctattc atatgctcta accttgagta cctatctatt 1800

ataataaaca agtatgtttt ataattattt tgatcttgat atacttggat gatggcatat 1860

gcagcagcta tatgtggatt tttttagccc tgccttcata cgctatttat ttgcttggta 1920

ctgtttcttt tgtcgatgct caccctgttg tttggtgtta cttctgcaga ccatggtgaa 1980

caaacacttc ttgtcccttt cggtcctcat cgtcctcctt ggcctctcct ccaacttgac 2040

agccgggga tccgtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt 2100

[0028]

cgagctggac ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga 2160

tgccacctac ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc 2220

ctggcccacc ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga 2280

ccacatgaag cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg 2340

caccatcttc ttcaaggacg acggcaacta caagacccgc gccgaggtga agttcgaggg 2400

cgacaccctg gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat 2460

cctggggcac aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa 2520

gcagaagaac ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt 2580

gcagctcgcc gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc 2640

cgacaaccac tacctgagca cccagtccgc cctgagcaaa gaccccaacg agaagcgcga 2700

tcacatggtc ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct 2760

gtacaagtaa gagctcaagc ttcgaattct gcagtcgacg gtaccgcggg cccgggatcc 2820

gcccctctcc ctcccccccc cctaacgtta ctggccgaag ccgcttggaa taaggccggt 2880

gtgcgtttgt ctatatgtta ttttccacca tattgccgtc ttttggcaat gtgagggccc 2940

ggaaacctgg ccctgtcttc ttgacgagca ttcctagggg tctttccccc ctcgccaaag 3000

gaatgcaagg tctgttgaat gtcgtgaagg aagcagttcc tctggaagct tcttgaagac 3060

aaacaacgtc tgtagcgacc ctttgcaggc agcggaaccc cccacctggc gacaggtgcc 3120

tctgcggcca aaagccacgt gtataagata cacctgcaaa ggcggcacaa ccccagtgcc 3180

acgttgtgag ttggatagtt gtggaaagag tcaaatggct ctcctcaagc gtattcaaca 3240

aggggctgaa ggatgcccag aaggtacccc attgtatggg atctgatctg gggcctcggt 3300

gcacatgctt tacatgtgtt tagtcgaggt taaaaaaacg tctaggcccc ccgaaccacg 3360

gggacgtggt tttcctttga aaaacacgat gataatatgg ccacctcgag accatgattg 3420

aacaagatgg attgcacgca ggttctccgg ccgcttgggt ggagaggcta ttcggctatg 3480

actgggcaca acagacaatc ggctgctctg atgccgccgt gttccggctg tcagcgcagg 3540

ggcgcccggt tcttttgtc aagaccgacc tgtccggtgc cctgaatgaa ctgcaggacg 3600

aggcagcgcg gctatcgtgg ctggccacga cgggcgttcc ttgcgcagct gtgctcgacg 3660

ttgtcactga agcgggaagg gactggctgc tattgggcga agtgccgggg caggatctcc 3720

tgtcatctca ccttgctcct gccgagaaag tatccatcat ggctgatgca atgcggcggc 3780

tgcatacgct tgatccggct acctgcccat tcgaccacca agcgaaacat cgcatcgagc 3840

gagcacgtac tcggatggaa gccggtcttg tcgatcagga tgatctggac gaagagcatc 3900

aggggctcgc gccagccgaa ctgttcgcca ggctcaaggc gcgcatgccc gacggcgatg 3960

atctcgtcgt gacccatggc gatgcctgct tgccgaatat catggtggaa aatggccgct 4020

[0029]

tttctggatt catcgactgt ggccggctgg gtgtggcgga ccgctatcag gacatagcgt 4080

tggctacccg tgatattgct gaagagcttg gcggcgaatg ggctgaccgc ttcctcgtgc 4140

tttacggtat cgccgctccc gattcgcagc gcatcgcctt ctatcgcctt cttgacgagt 4200

tcttctgagc gggactctgg ggttcgaaat gaccgaccaa gcgacgccca acctgccatc 4260

acgagatttc gattccaccg ccgccttcta tgaaaggttg ggcttcggaa tcgttttccg 4320

ggacgccggc tggatgatcc tccagcgcgg ggatctcatg ctggagttct tcgcccacgg 4380

gatctctgcg gaacaggcgg tcgaaggtgc cgatatcatt acgacagcaa cggccgacaa 4440

gcacaacgcc acgatcctga gcgacaatat gatcgggccc ggcgtccaca tcaacggcgt 4500

cggcggcgac tgcccaggca agaccgagat gcaccgcgat atcttgctgc gttcggatat 4560

tttcgtggag ttcccgccac agacccggat gatccccgat cgttcaaaca tttggcaata 4620

aagtttctta agattgaatc ctgttgccgg tcttgcgatg attatcatat aatttctgtt 4680

gaattacgtt aagcatgtaa taattaacat gtaatgcatg acgttattta tgagatgggt 4740

ttttatgatt agagtcccgc aattatacat ttaatacgcg atagaaaaca aaatatagcg 4800

cgcaaactag gataaattat cgcgcgcggt gtcatctatg ttactagatc gggcctcctg 4860

tcaatgctgg cggcggctct ggtctaga 4888

<210> 44
<211> 1963
<212> DNA
<213> 人工序列

<220>
<223> 合成寡核苷酸

<400> 44
tgatccccta atgagcattg catgtctaag ttataaaaaa ttaccacata tttttttgtc 60

acacttgttt gaagtgcagt ttatctatct ttatacatat atttaaactt cactctacga 120

ataatataat ctatagtact acaataatat cagtgtttta gagaatcata taaatgaaca 180

gttagacatg gtctaaagga caattgagta ctttgacaac aggactctac agttttatct 240

ctttagtgtg catgtgttct cctttttttt tgcaaatagc ttcacctata taatacttca 300

tccattttat tagtacatcc atttagggtt tagggttaat ggtttttata gactaatttt 360

tttagtacat ctattttatt ctattttagc ctctaaatta agaaaactaa aactctattt 420

tagttttttt atttaataat ttagatataa aatagaataa aataaagtga ctaaaaatta 480

aacaaatacc ctttaagaaa ttaaaaaaac taaggaaaca tttttcttgt ttcgagtaga 540

taatgccagc ctgttaaacg ccgtcgacga gtctaacgga caccaaccag cgaaccagca 600

gcgtcgcgtc gggccaagcg aagcagacgg cacggcatct ctgtcgctgc ctctgggccc 660

ctctcgagag ttccgctcca ccgttggact tgctccgctg tcggcatcca gaaattgcgt 720

ggcggagcgg cagacgtgag ccggcacggc aggcggcctc ctcctcctct cacggcaccg 780

[0030]

gcagctacgg gggattcctt tcccaccgct ccttcgcttt cccttcctcg cccgccgtaa 840

taaatagaca ccccctccac accctctttc cccaacctcg tgttgttcgg agcgcacaca 900

cacacaacca gatctccccc aaatccaccc gtcggcacct ccgcttcaag gtacgccgct 960

cgtcctcccc cccccccctct ctaccttctc tagatcggcg ttccggtcca tggttagggc 1020

ccggtagttc tacttctgtt catgtttgtg ttagatccgt gtttgtgata gatccgtgct 1080

gctagcgttc gtacacggat gcgacctgta cgtcagacac gttctgattg ctaacttgcc 1140

agtgtttctc tttggggaat cctgggatgg ctctagccgt tccgcagacg ggatcgattt 1200

catgattttt ttttgtttcg ttgcataggg tttggtttgc ccttttcctt tatttcaata 1260

tatgccgtgc acttgtttgt cgggtcatct tttcatgttt tttttttggc ttggttgtga 1320

tgatgtggtc tggttgggcg gtcgttctag atcggagtag aattctgttt caaactacct 1380

ggtggattta ttaattttgg atctgtatgt gtgtgccata catattcata gttacgaatt 1440

gaagatgatg gatggaaata tcgatctagg ataggtatac atgttgatgc gggttttact 1500

gatgcatata cagagatgct ttttgttcgc ttggttgtga tgatgtggtg tggttgggcg 1560

gtcgttcatt cgttctagat cggagtagaa tactgtttca aactaactgg tggatttatt 1620

aattttggat ctgtatgtgt gtgccataca tcttcatagt tacgagttta agatgatgga 1680

tggaagtatc gatctaggat aggtatacat gttgatgtgg gttttactga tgcatataca 1740

tgatggcata tgcagcatct attcatatgc tctaaccttg agtacctatc tattataata 1800

aacaagtatg ttttataatt attttgatct tgatatactt ggatgatggc atatgcagca 1860

gctatatgtg gatttttttа gccctgcctt catacgctat ttatttgctt ggtactgttt 1920

cttttgtcga tgctcaccct gttgtttggt gttacttctg cag 1963

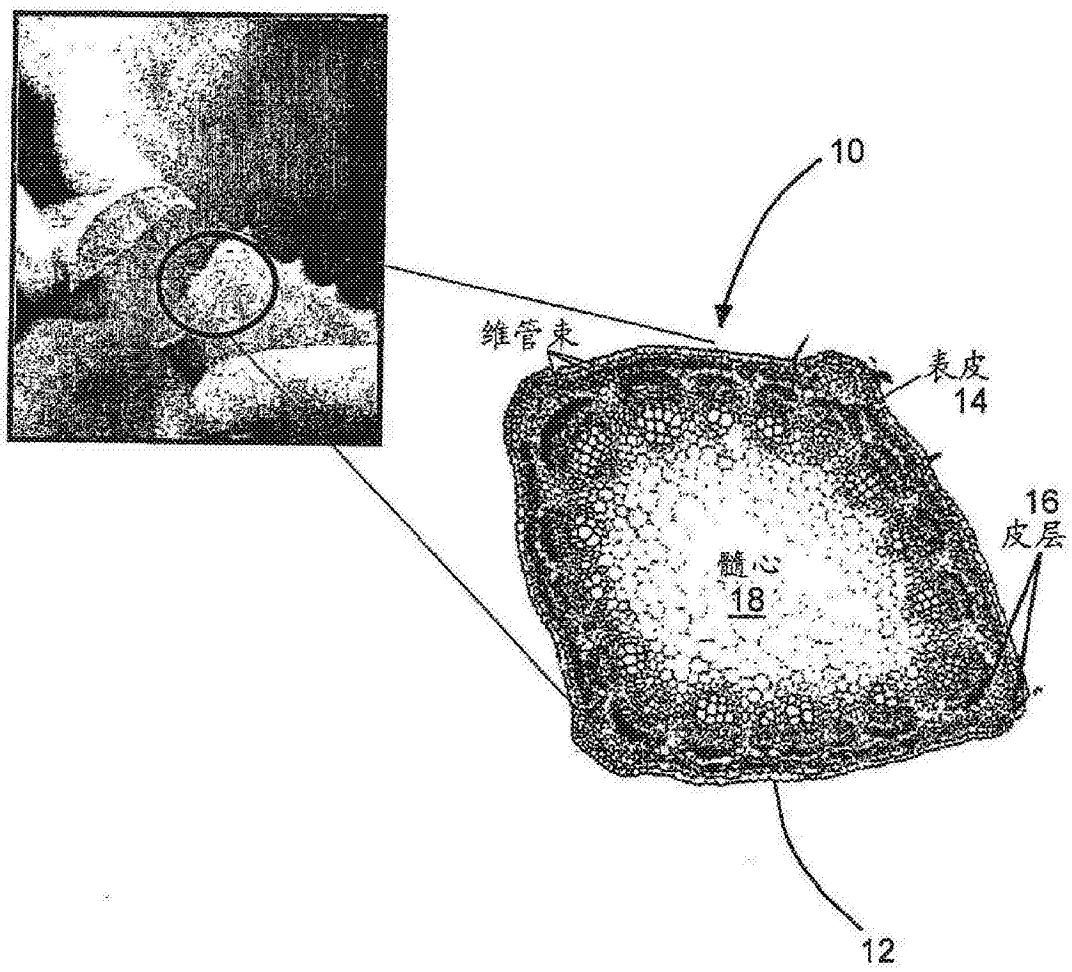


图 1

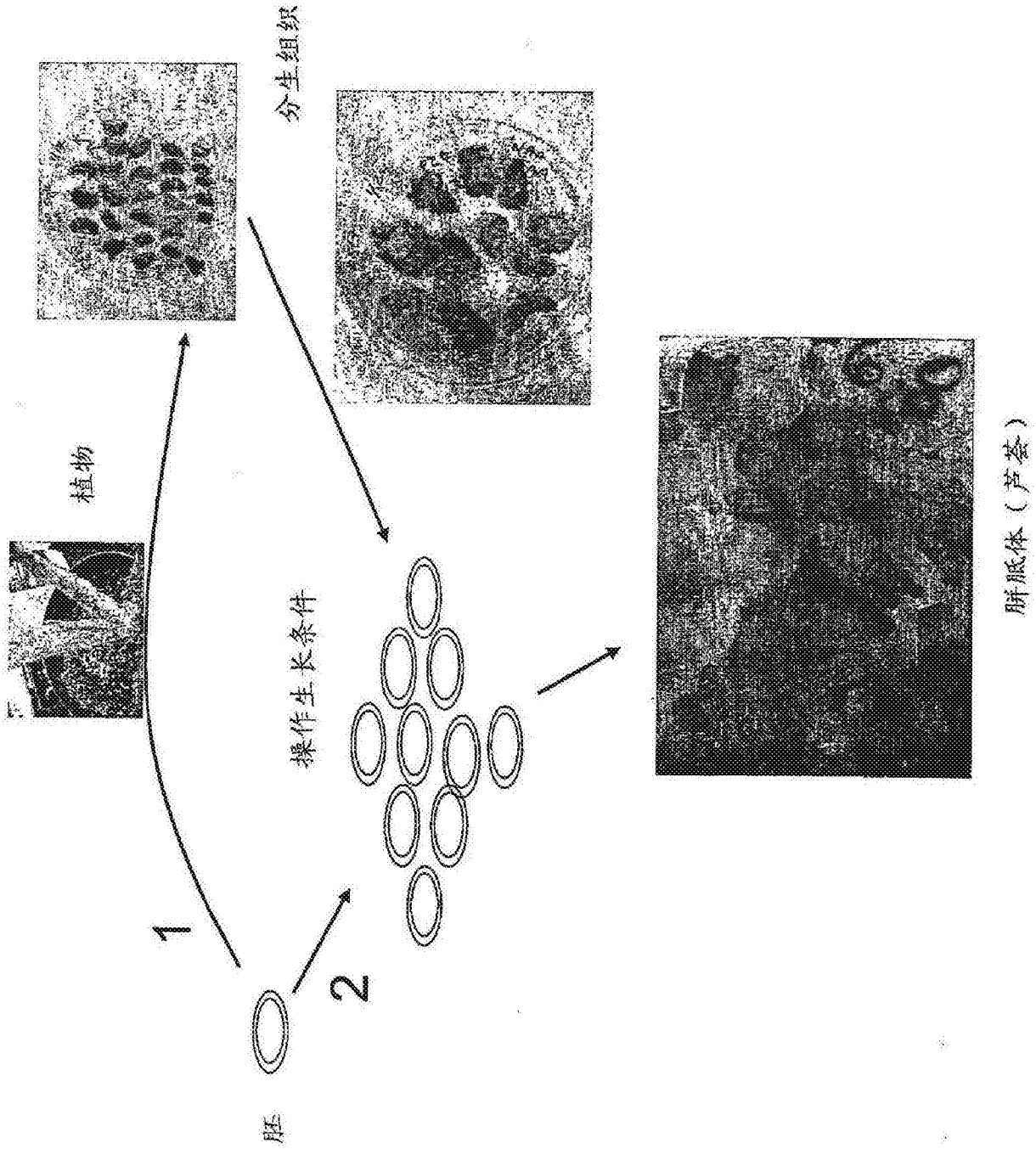


图 2

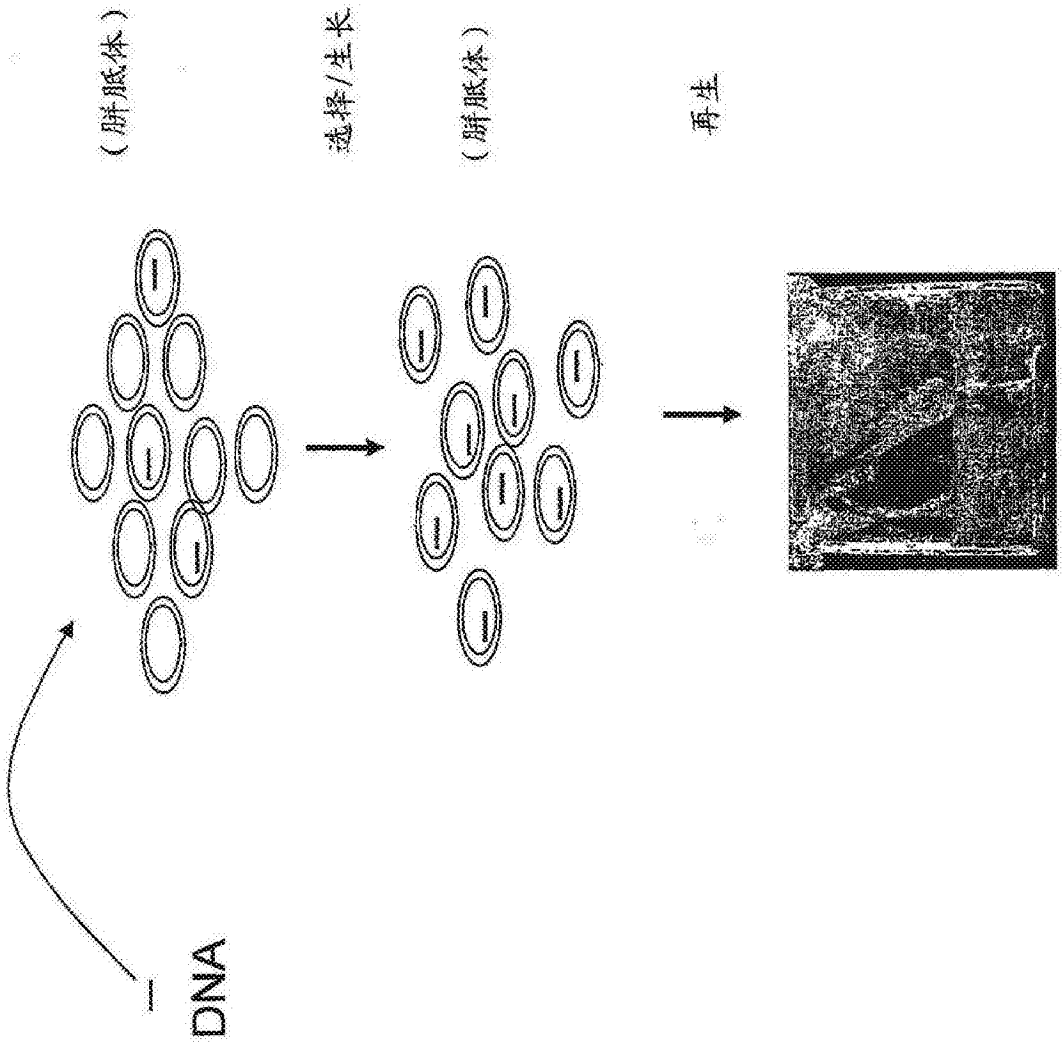


图 3

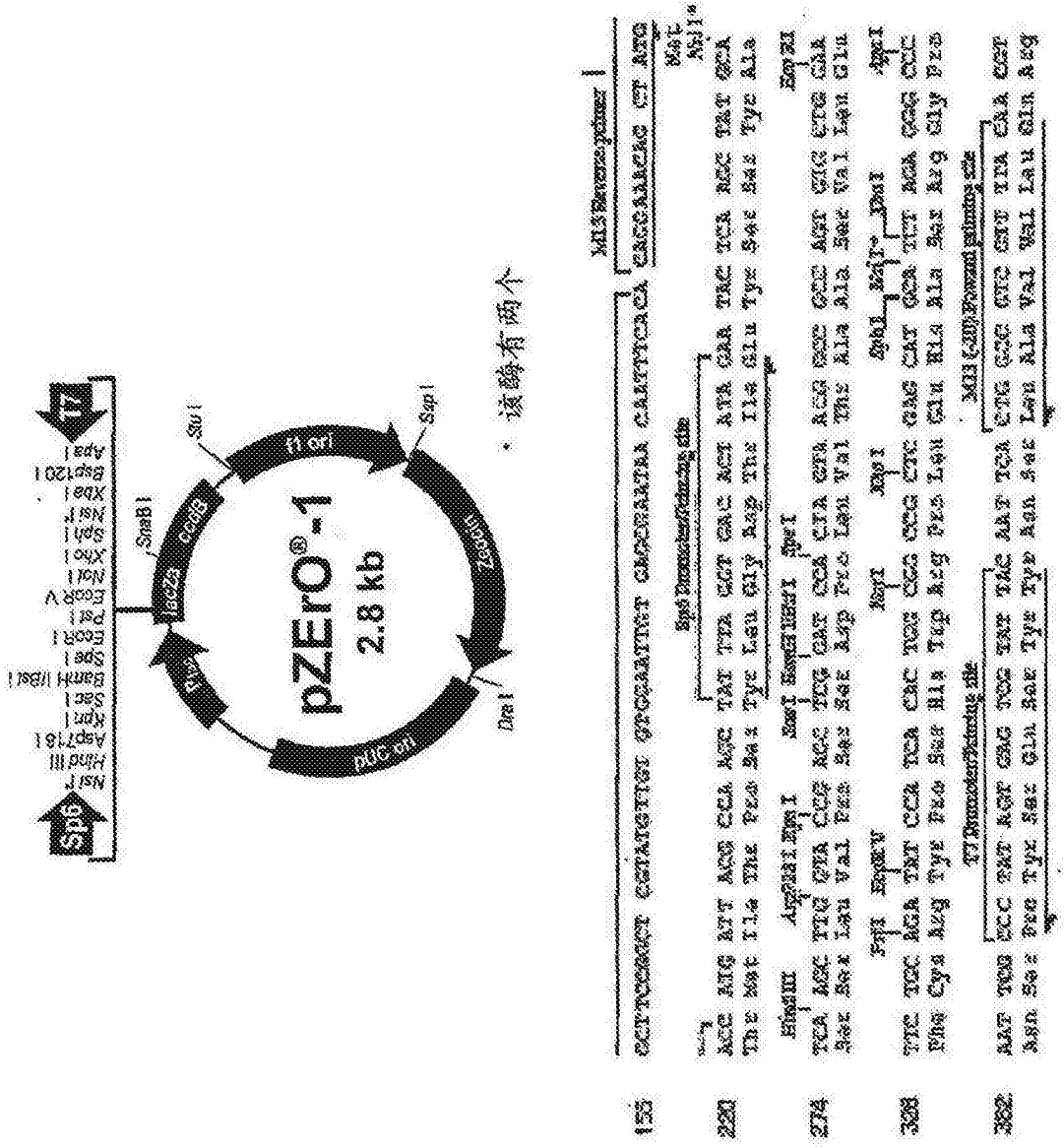

pZErO®-1
2.8 kb
f1 ori
Sap I
zeocin
pUC ori
Dra I
Sbf I
ccdB
lacZ
SnaB I
Sfo I
T7
Sp6
Nsi I
Hind III
Asp718 I
Kpn I
Sac I
BamH I/Bst I
Spe I
EcoR I
Pst I
EcoR V
Nco I
Xho I
Nsi I
Sph I
Xba I
Bsp120 I
Apa I
· 该酶有两个

图 4

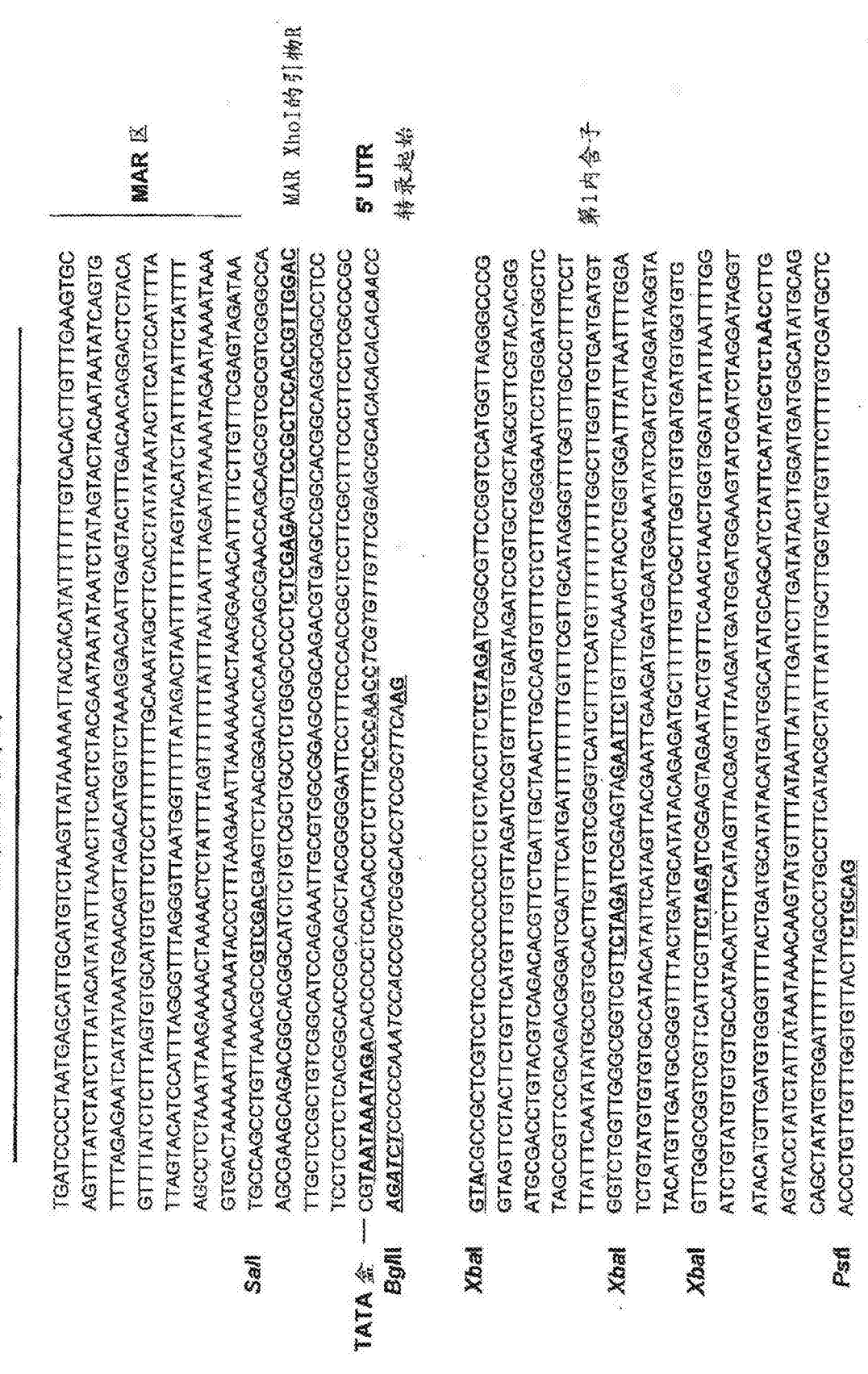

泛素启动子序列
TGATCCCCTAATGAGCATTGCATGTCTAAGTTATAAAAAATTACCACATATTTTTTTGTCACACTTGTTTGAAGTGC
AGTTTATCTATCTTTATACATATATTTAAACTTCACTCTACGAATAATATAATCTATAGTACTACAATAATATCAGTG
TTTTAGAGAATCATATAAATGAACAGTTAGACATGGTCTAAAGGACAATTGAGTACTTTGACAACAGGACTCTACA
GTTTTATCTCTTTAGTGTGCATGTGTTCTCCTTTTTTTTTGCAAATAGCTTCACCTATATAATACTTCATCCATTTTA
TTAGTACATCCATTTAGGGTTTAGGGTTAATGGTTTTTATAGACTAATTTTTTTTAGTACATCTATTTTATTCTATTTT
AGCCTCTAAATTAAGAAAACTAAAACTCTATTTTAGTTTTTTTATTTAATAATTTAGATATAAAATAGAATAAAATAAA
GTGACTAAAAAATTAAACAAATACCCTTTAAGAAATTAAAAAAACTAAGGAAACATTTTTCTTGTTTCGAGTAGATAA
SalI
TGCCAGCCTGTTAAACGCCGTCGACGAGTCTAACGGACACCAACCAGCGAACCAGCAGCGTCGCGTCGGGCCA
AGCGAAGCAGACGGCACGGCATCTCTGTCGCTGCCTCTGGCCCCTCTCGAGAGTTCCGCTCCACCGTTGGAC
MAR Xho1的引物R
TTGCTCCGCTGTCGGCATCCAGAAATTGCGTGGCGGAGCGGCAGACGTGAGCCGGCACGGCAGGCGGCCTCC
TCGTGCTCTCACGGCACCGGCAGCTACGGGGGATTCCTTTCCCACCGCTCCTTCGCTTTCCCTTCCTCGCCGCC
TATA盒 — CGTAATAAATAGACACCCCCTCCACACCCTCTTTCCCCAACCTCGTGTTGTTCGGAGCGCACACACACACAACCC
BglII
AGATCTCCCCCAAATCCACCCGTCGGCACCTCCGCTTCAAG
5' UTR
转录起始
XbaI
GTACGCCGCTCGTCCTCCCCCCCCCCCTCTCTACCTTCTCTAGATCGGCGTTCCGGTCCATGGTTAGGGCCCG
GTAGTTCTACTTCTGTTCATGTTTGTGTTAGATCCGTGTTTGTGTTAGATCCGTGCTAGCGTTCGTACACGG
ATGCGACCTGTACGTCAGACACGTTCTGATTGCTAACTTGCCAGTGTTTCTCTTTGGGGAATCCTGGGATGGCTC
TAGCCGTTCCGCAGACGGGATCGATTTCATGATTTTTTTTGTTTCGTTGCATAGGGTTTGGTTTGCCCTTTTCCT
TTATTTCAATATATGCCGTGCACTTGTTTGTCGGTCATCTTTTCATGTTTTTTTTTGTCTTGGTTGTGATGATGT
XbaI
GGTCTGGTTGGCGGTCGTTCTAGATCCGGAGTAGAATTCTGTTTCAAACTACCTGGTGGATTTATTAATTTTGGA
TCTGTATGTGTGTGCCATACATATTCATAGTTACGAATTGAAGATGATGGAAATATCGATCTAGGATAGGTA
TACATGTTGATGCGGGTTTTACTGATGCATATACAGAGATGCTTTTTGTTCGCTTGGTTGTGATGATGTGGTGTG
XbaI
GTTGGGCGGTCGTTCATTCGTTCTAGATCGGAGTAGAATACTGTTTCAAACTAACTGGTGGATTTATTAATTTTGG
ATCTGTATGTGTGTGCCATACATCTTCATAGTTACGAGTTTAAGATGATGGATGGAAGTATCGATCTAGGATAGGT
ATACATGTTGATGTGGGTTTTACTGATGCATATACATGATGGCATATGCAGCATCTATTCATATGCTCTAACCTTG
AGTACCTATCTATTATAATAAACAAGTATGTTTTATAATTATTTTGATCTTGATATACTTGGATGATGGCATATGCAG
CAGCTATATGTGGATTTTTTTAGCCCTGCCTTCATACGCTATTTATTTGCTTGGTACTGTTTCTTTTGTCGATGCTC
PstI
ACCCTGTTGTTTGGTGTTACTTCTGCAG

图 5

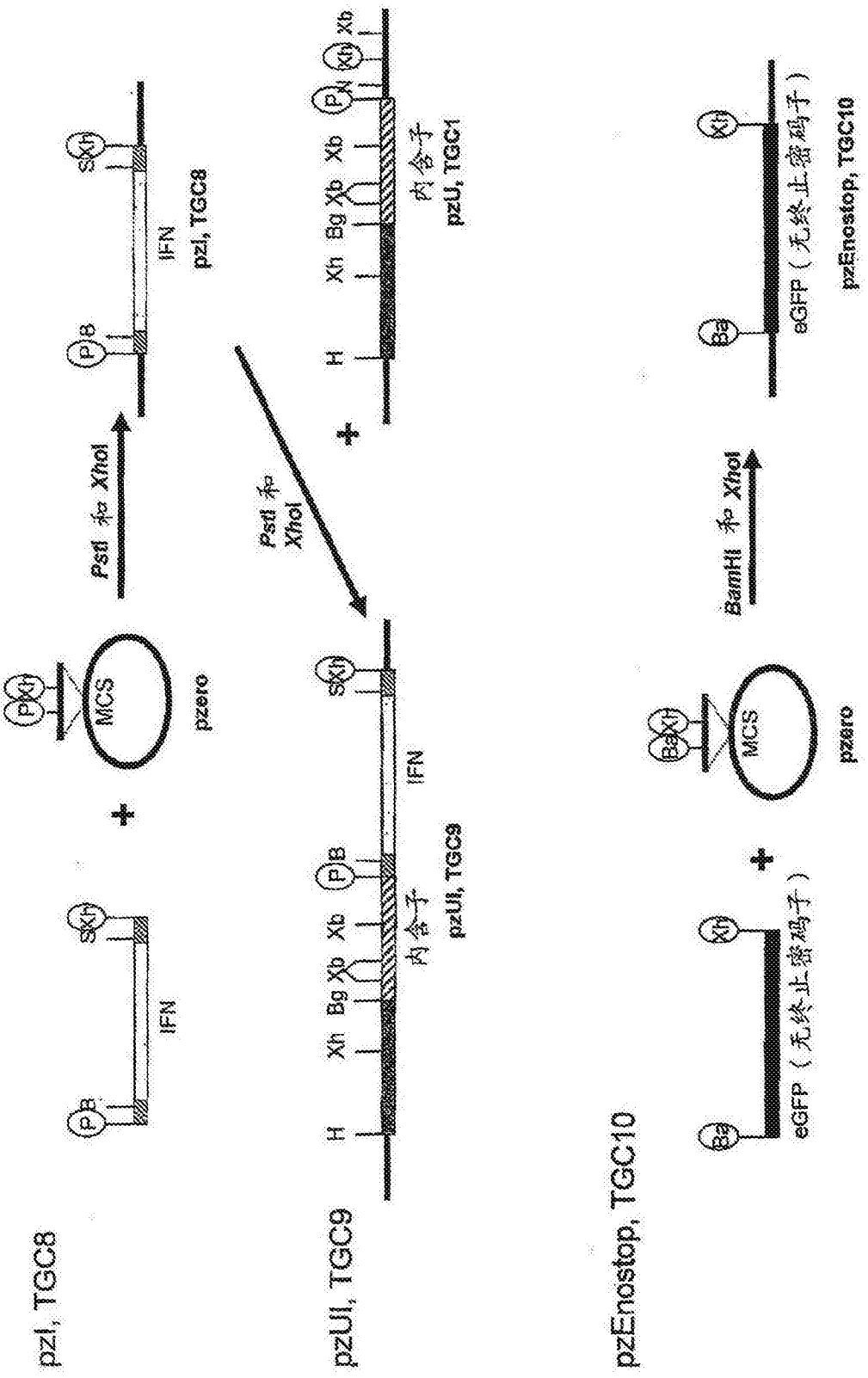

pzI, TGC8
pzero
pzI, TGC8
pzUI, TGC9
pzU, TGC1
pzUI, TGC9
pzEnostop, TGC10
pzero
pzEnostop, TGC10
IFN
MCS
内含子
eGFP（无终止密码子）
PstI 和 XhoI
BamHI 和 XhoI

图 6A

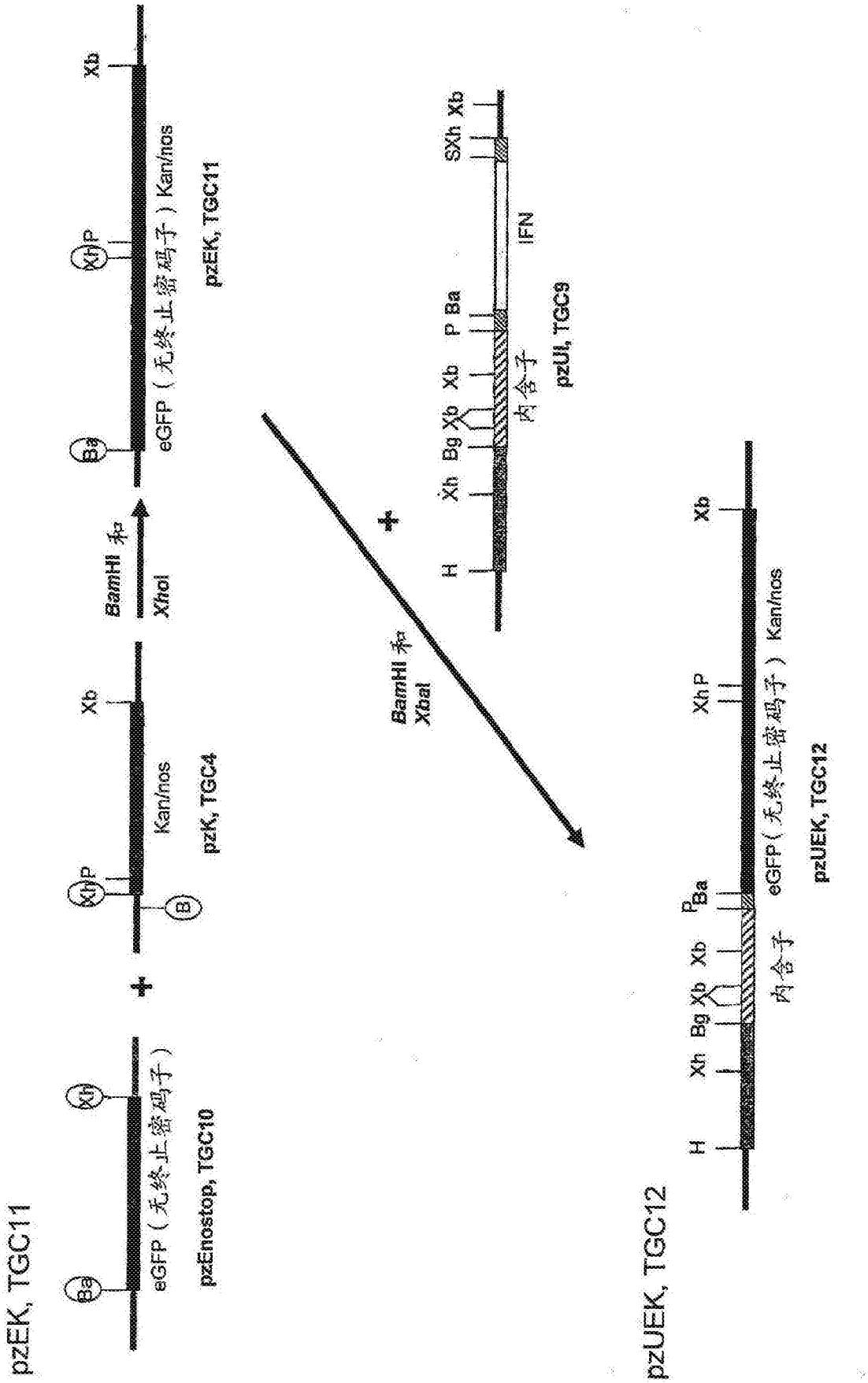

pzEK, TGC11
Ba
Xh
eGFP（无终止密码子）
pzEnostop, TGC10
Xh P
Xb
Kan/nos
B
pzK, TGC4
BamHI 和
XhoI
Ba
Xh P
Xb
eGFP（无终止密码子）Kan/nos
pzEK, TGC11
BamHI 和
XbaI
H
Xh
Bg
Xb
Xb
P Ba
内含子
IFN
SXh Xb
pzUI, TGC9
pzUEK, TGC12
H
Xh
Bg
Xb
Xb
P Ba
内含子
eGFP（无终止密码子）Kan/nos
Xh P
Xb
pzUEK, TGC12

图 6B

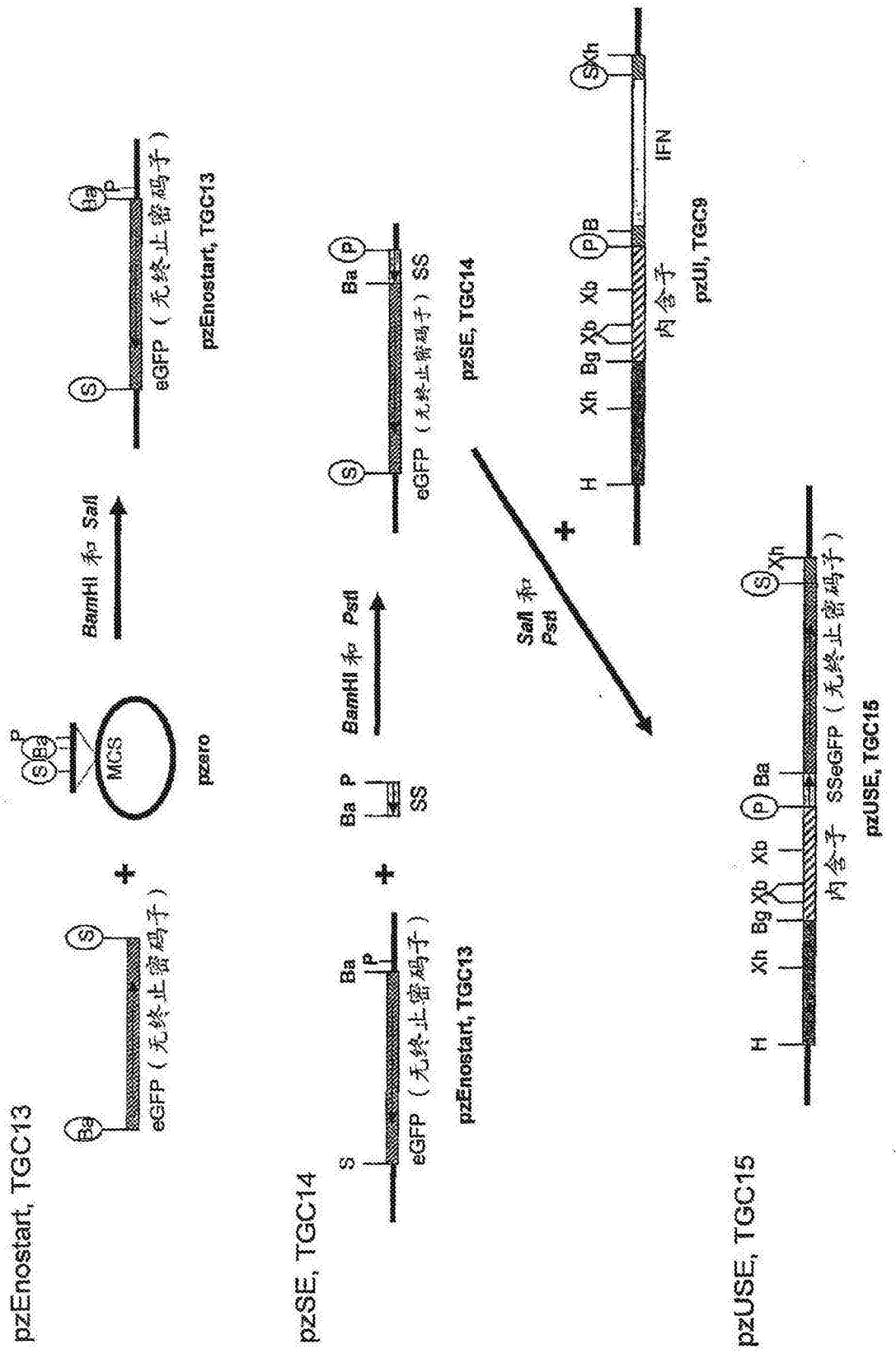

pzEnostart, TGC13
Ba
S
eGFP（无终止密码子）
+
S
Ba
P
MCS
pzero
BamHI 和 SalI
S
Ba
P
eGFP（无终止密码子）
pzEnostart, TGC13
pzSE, TGC14
S
Ba
P
eGFP（无终止密码子）
+
Ba P
SS
pzEnostart, TGC13
BamHI 和 PstI
S
Ba
P
eGFP（无终止密码子）SS
pzSE, TGC14
SalI 和
PstI
+
H
Xh
Bg
Xb
Xb
Ba
P
内含子
IFN
Sal
Xh
pzUI, TGC9
pzUSE, TGC15
H
Xh
Bg
Xb
Xb
Ba
P
内含子 SSeGFP（无终止密码子）
S
Xh
pzUSE, TGC15

图 7A

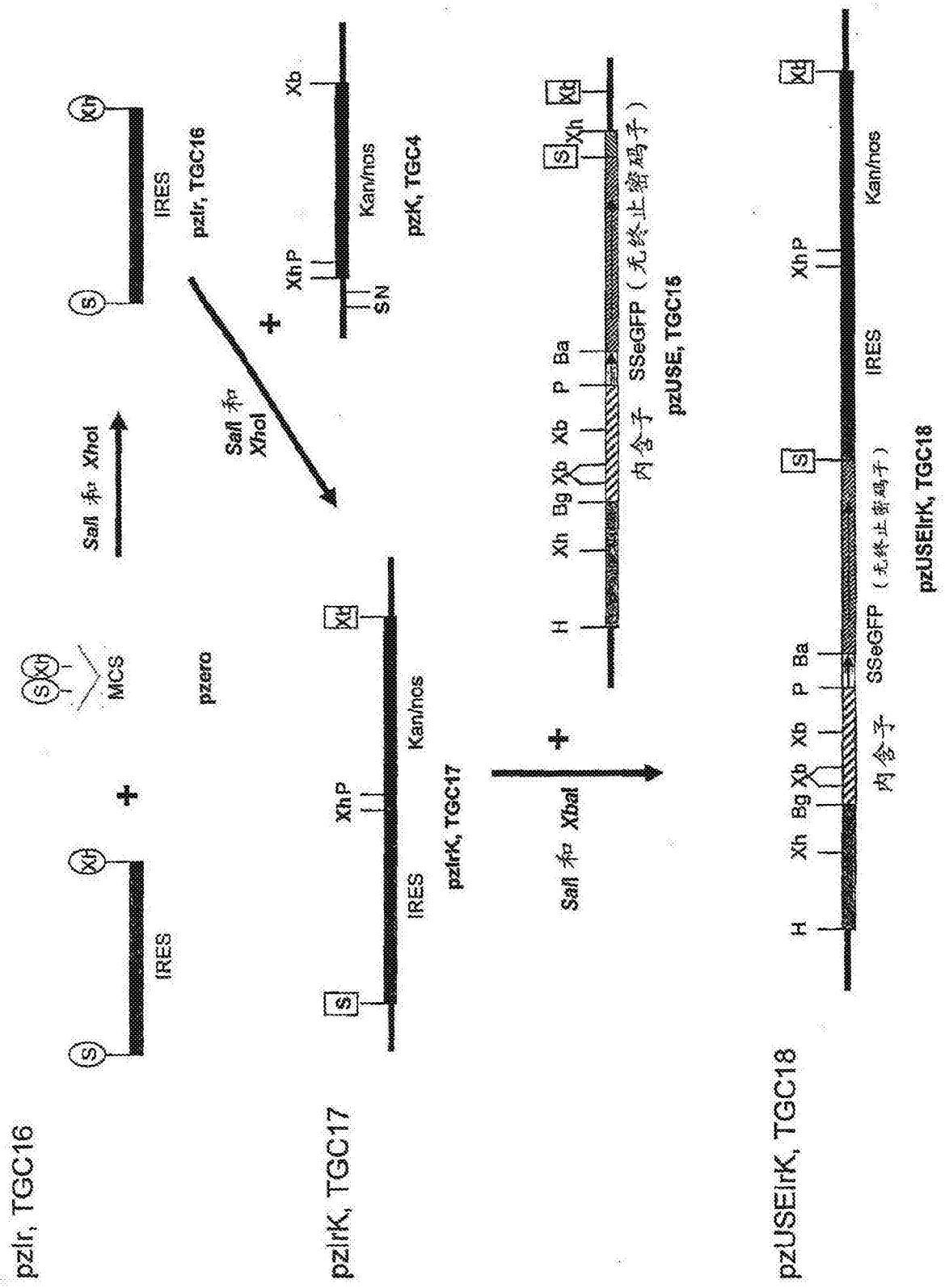

pzlr, TGC16
IRES
pzero
MCS
SalI 和 XhoI
pzlr, TGC16
pzlrK, TGC17
Kan/nos
pzK, TGC4
SN
pzlrK, TGC17
pzUSE, TGC15
SSeGFP（无终止密码子）
内含子
SalI 和 XbaI
pzUSElrK, TGC18

图 7B

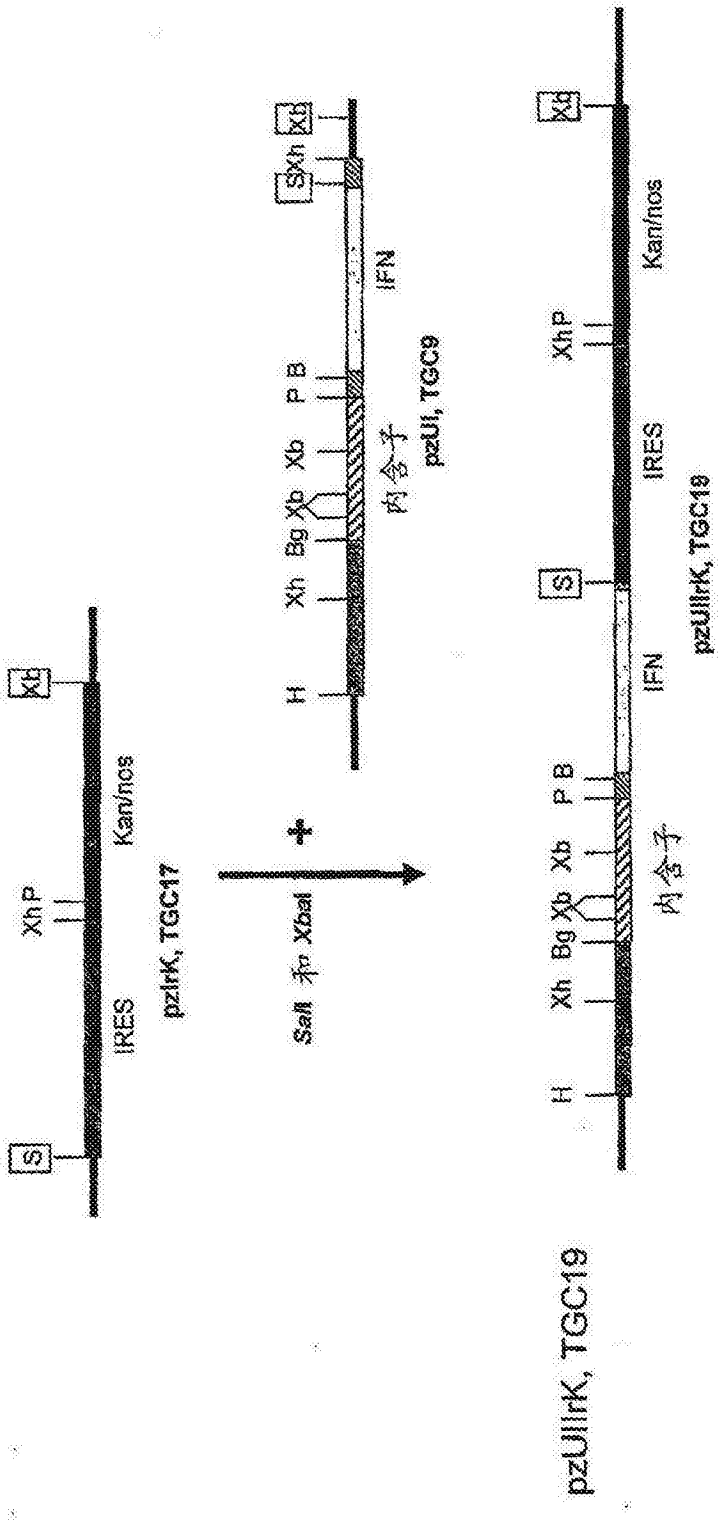

Xb
Kan/nos
XhP
IRES
S
pzIrK, TGC17
+
SalI 和 XbaI
Sal Xh
IFN
P B
Xb
Xb
Bg
Xh
H
内含子
pzUI, TGC9
pzUIIrK, TGC19
pzUIIrK, TGC18

图 7C

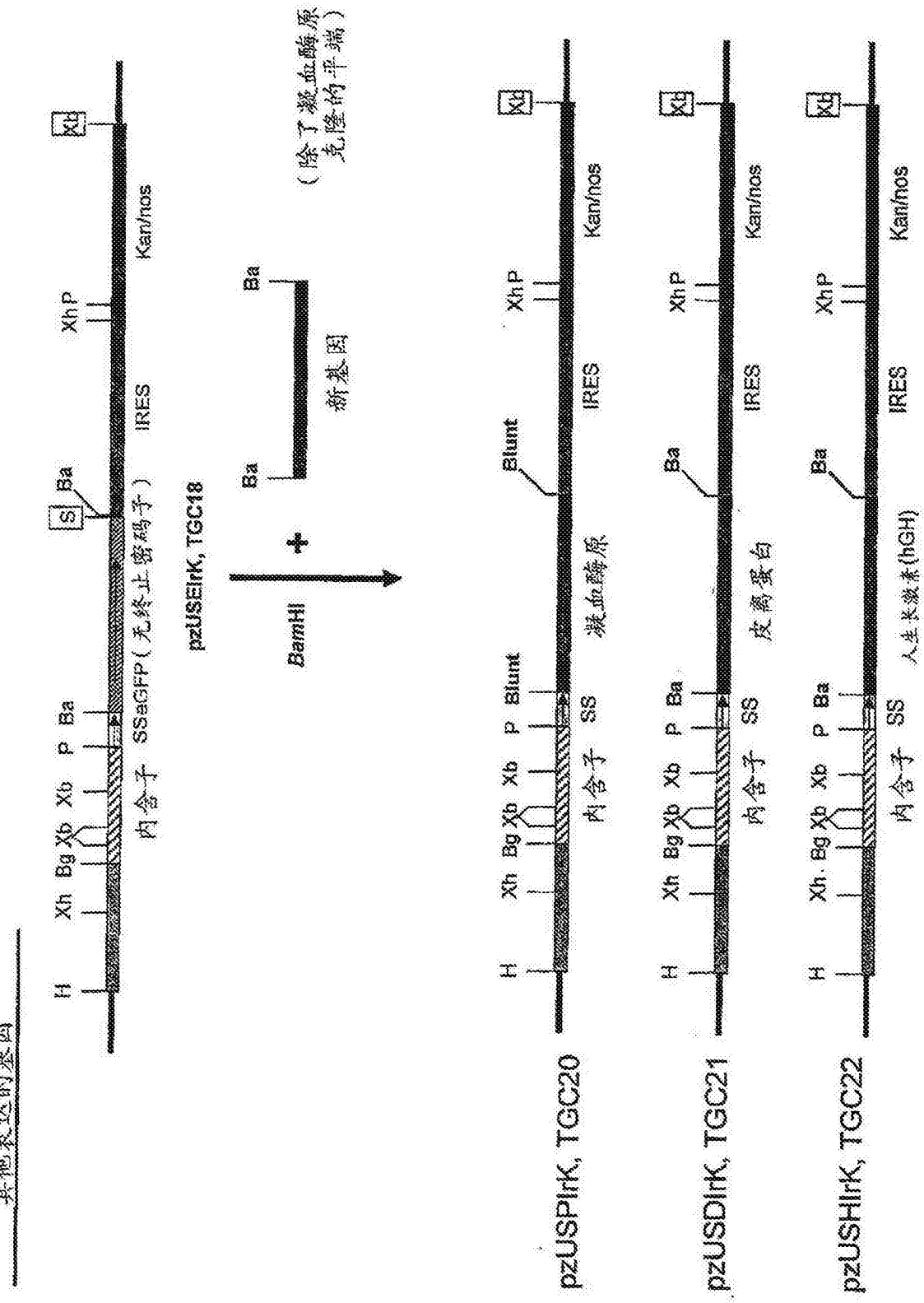
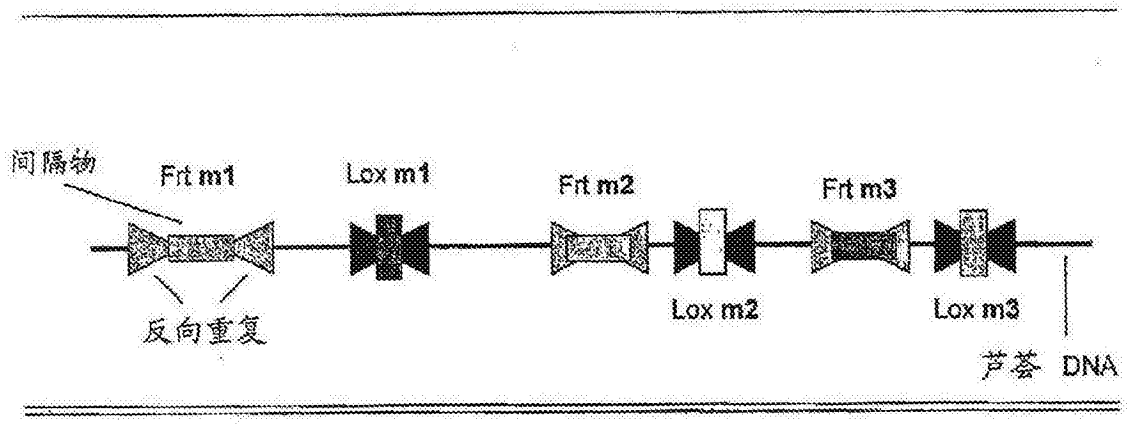

其他表达的基因
H
Xh
Bg
Xb
P
Ba
内含子
SSaGFP(无终止密码子)
IRES
XhP
Kan/nos
pzUSEIrK, TGC18
BamHI
新基因
(除了凝血酶原无隆的平端)
Blunt
SS
凝血酶原
pzUSPIrK, TGC20
皮肤蛋白
pzUSDIrK, TGC21
人生长激素(hGH)
pzUSHIrK, TGC22

图 8


间隔物
Frt m1
Lox m1
Frt m2
Frt m3
反向重复
Lox m2
Lox m3
芦荟 DNA

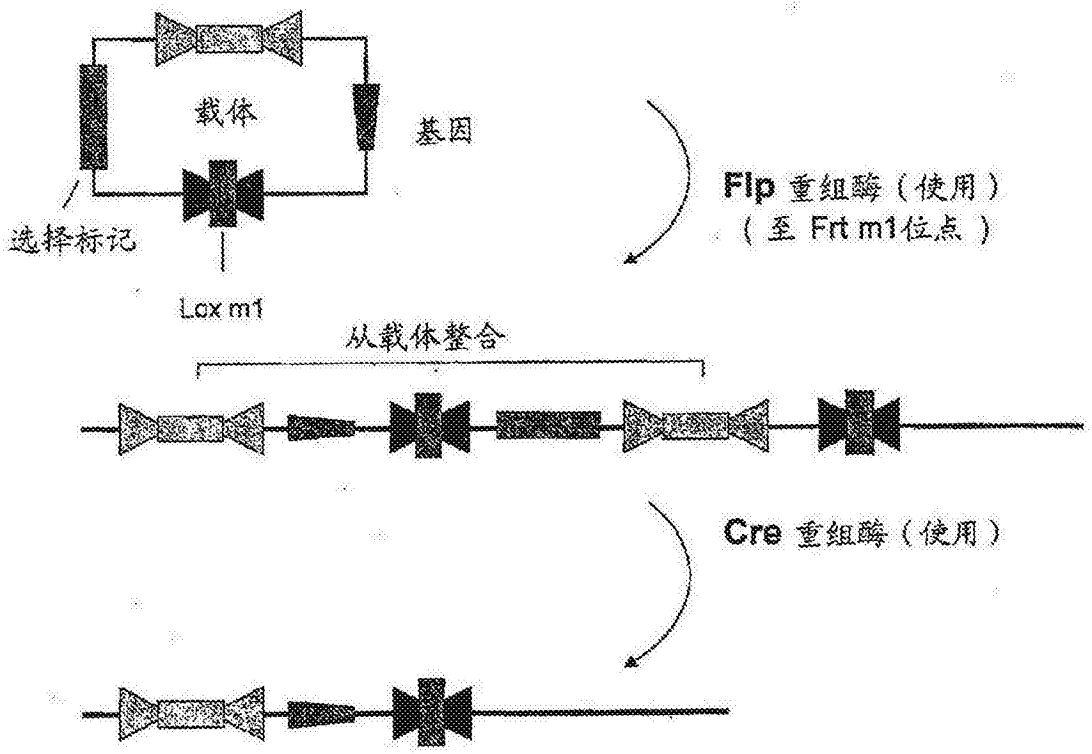

载体
基因
选择标记
Lox m1
Flp 重组酶（使用）
（至 Frt m1位点）
从载体整合
Cre 重组酶（使用）

图 9