

[19] 中华人民共和国国家知识产权局



# [12] 发明专利说明书

专利号 ZL 99806120.4

[51] Int. Cl.

*C12N 15/31 (2006.01)*

*C12N 15/53 (2006.01)*

*C12N 15/55 (2006.01)*

*C12N 15/60 (2006.01)*

*C12P 17/18 (2006.01)*

*C12N 9/04 (2006.01)*

[45] 授权公告日 2007 年 1 月 31 日

[11] 授权公告号 CN 1297662C

[51] Int. Cl. (续)

*C12N 9/16 (2006.01)*

*C12N 9/88 (2006.01)*

*C12N 1/21 (2006.01)*

*C07K 14/24 (2006.01)*

[22] 申请日 1999.5.12 [21] 申请号 99806120.4

[30] 优先权

[32] 1998. 5. 12 [33] US [31] 60/085,190

[86] 国际申请 PCT/US1999/010356 1999.5.12

[87] 国际公布 WO1999/058686 英 1999.11.18

[85] 进入国家阶段日期 2000.11.13

[73] 专利权人 纳慕尔杜邦公司

地址 美国特拉华州威尔明顿

共同专利权人 詹伦卡国际有限公司

[72] 发明人 G·M·怀特德 B·布尔图伊斯

D·E·特林布尔 A·A·加滕拜

[56] 参考文献

WO9635795A1 1996.11.14 C12N15/60

审查员 马秋娟

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 卢新华 谭明胜

权利要求书 3 页 说明书 59 页

[54] 发明名称

用含有维生素 B<sub>12</sub> 转运基因的重组生物体生产  
1,3-丙二醇的方法

[57] 摘要

提供了含有编码甘油脱水酶, 1, 3 - 丙二醇基因, 1, 3 - 丙二醇氧化还原酶基因, 编码维生素 B<sub>12</sub> 受体前体 (BtuB) 的基因, 编码维生素 B<sub>12</sub> 转运系统通透酶蛋白 (BtuC) 的基因和编码维生素 B<sub>12</sub> 转运 ATP - 结合蛋白 (BtuD) 的基因。使重组微生物体和碳源底物接触并且从生长培养基中分离 1, 3 - 丙二醇。

1. 一种 1,3-丙二醇的生物生产方法, 包括:
  - (i)将转化的宿主细胞与至少一种可发酵碳源以及有效剂量的维生素 B<sub>12</sub> 接触从而生产 1,3-丙二醇, 转化的宿主细胞含有:
    - (a)至少一个拷贝的基因, 该基因编码一种具有脱水酶活性的蛋白;
    - (b)至少一个拷贝的基因, 该基因编码一种具有氧化还原酶活性的蛋白;
    - (c)至少一个拷贝的基因, 该基因编码一种维生素 B<sub>12</sub> 受体前体蛋白;
    - (d)至少一个拷贝的基因, 该基因编码一种维生素 B<sub>12</sub> 转运系统通透酶蛋白; 和
    - (e)至少一个拷贝的基因, 该基因编码一种维生素 B<sub>12</sub> 转运 ATP - 或 GTP - 结合蛋白; 并且
  - (ii)回收从步骤(i)产生的 1, 3 - 丙二醇。
2. 权利要求 1 的方法, 其中步骤(i)(a)中编码具有脱水酶活性的蛋白的基因编码的酶选自甘油脱水酶和二元醇脱水酶。
3. 权利要求 1 的方法, 其中步骤(i)(a)和(i)(b)的基因分别独立地分离自克雷伯氏菌类、柠檬酸杆菌类、沙门氏菌类或梭菌类。
4. 权利要求 1 的方法, 其中步骤(i)(c)、(i)(d)和(i)(e)中的基因分别独立地分离自埃希氏菌类、沙门氏菌类、克雷伯氏菌类或柠檬酸杆菌类。
5. 权利要求 1 中方法, 其中
  - (1) 步骤(i)(c)的基因是选自 SEQ ID NO: 1 和 SEQ ID NO: 2 的 btuB 基因;
  - (2) 步骤(i)(d)的基因是 SEQ ID NO: 3 的 btuC 基因;和
  - (3) 步骤(i)(e)的基因是 SEQ ID NO: 4 的 btuD 基因。
6. 权利要求 1 的方法, 其中可发酵碳源选自可发酵碳水化合物、单碳底物和它们的混合物。
7. 权利要求 1 的方法, 其中可发酵碳源选自单糖类、寡糖类、多糖类、单碳底物类、甘油、二羟基丙酮和含碳胺类。
8. 权利要求 1 中方法, 其中转化的宿主细胞还含有至少一个拷贝的编码甘油 - 3 - 磷酸脱氢酶的基因以及至少一个拷贝的编码甘油 - 3 - 磷

酸酶的基因。

9. 权利要求 1 中方法，其中宿主细胞选自细菌、酵母和丝状真菌。

10. 权利要求 9 中方法，其中宿主细胞选自如下的属：柠檬酸杆菌属、肠杆菌属、梭菌属、克雷伯氏菌属、气杆菌属、乳杆菌属、曲霉属、糖酵母属、裂殖糖酵母属、结合糖酵母属、毕赤酵母属、克鲁维氏酵母属、假丝酵母属、汉逊氏酵母属、德巴利氏酵母、毛霉属、球拟酵母属、甲基杆菌属、埃希氏菌属、沙门氏菌属、芽孢杆菌属、链霉菌属和假单胞菌属。

11. 权利要求 1 中方法，其中，以摩尔比计，有效剂量的维生素 B<sub>12</sub> 为脱水酶的 0.1-10 倍。

12. 一种转化的宿主细胞，含有：

(a) 至少一个拷贝的基因，该基因编码一种具有脱水酶活性的蛋白；

(b) 至少一个拷贝的基因，该基因编码一种具有氧化还原酶活性的蛋白；

(c) 至少一个拷贝的基因，该基因编码一种维生素 B<sub>12</sub> 受体前体蛋白；

(d) 至少一个拷贝的基因，该基因编码一种维生素 B<sub>12</sub> 转运系统通透酶蛋白；和

(e) 至少一个拷贝的基因，该基因编码一种维生素 B<sub>12</sub> 转运 ATP - 或 GTP - 结合蛋白。

13. 一种 1,3-丙二醇的生物生产方法，包括：

① 将转化的宿主细胞与 (a) 至少一种选自单糖类、寡糖类、多糖类、单碳底物、甘油、二羟基丙酮和含碳胺类的可发酵碳源以及 (b) 有效剂量的维生素 B<sub>12</sub> 接触从而生产 1,3-丙二醇，转化的宿主细胞含有：

(a) 至少一个拷贝的基因，该基因编码一种具有脱水酶活性的蛋白；

(b) 至少一个拷贝的基因，该基因编码一种具有氧化还原酶活性的蛋白；

(c) 至少一个拷贝的基因，该基因编码一种维生素 B<sub>12</sub> 受体前体蛋白；

(d) 至少一个拷贝的基因，该基因编码一种维生素 B<sub>12</sub> 转运系统通透酶蛋白；和

(e) 至少一个拷贝的基因，该基因编码一种维生素 B<sub>12</sub> 转运 ATP - 或 GTP - 结合蛋白；

(f) 至少一个拷贝的基因，该基因编码一种具有甘油 - 3 - 磷酸脱氢酶活性的蛋白；和

(g) 至少一个拷贝的基因，编码一种具有甘油 - 3 - 磷酸酶活性的蛋白，并且

②回收从步骤①产生的 1, 3 - 丙二醇。

用含有维生素 B<sub>12</sub> 转运基因的重组  
生物体生产 1,3-丙二醇的方法

### 发明领域

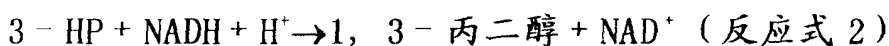
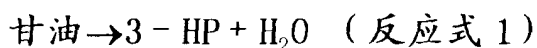
本发明涉及分子生物学领域以及使用重组生物体生产 1,3-丙二醇。更具体而言它描述了影响细胞内维生素 B<sub>12</sub> 的转运和有效地将碳源底物转化为 1,3-丙二醇的基因的克隆表达。

### 发明背景

1,3-丙二醇是一种应用于生产聚酯纤维以及制造聚氨基甲酸酯和环状化合物的单体。

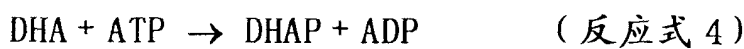
许多合成 1,3-丙二醇的化学途径是已知的。例如,可以这样制备 1,3-丙二醇,1) 在磷、水、一氧化碳、氢气和酸存在的情况下在催化剂上由环氧乙烷制备得到; 2) 通过催化丙烯醛的溶液相氢化,然后还原; 或 3) 由碳氢化合物如甘油在一氧化碳和氢气存在的情况下在含有元素周期表上第 VIII 组原子的催化剂上进行反应。尽管可能通过这些方法产生 1,3-丙二醇,但是这些方法费用昂贵,并且产生含有环境污染物质的废液。

一个多世纪以来已知可通过甘油的发酵生产 1,3-丙二醇。已经在细菌,例如柠檬酸杆菌属 (*Citrobacter*), 梭菌属 (*Clostridium*), 肠杆菌属 (*Enterobacter*), 泥杆菌属 (*Ilyobacter*), 克雷伯氏菌属 (*Klebsiella*), 乳杆菌属 (*Lactobacillus*), 和暗杆菌属 (*Pelobacter*) 中发现了能产生 1,3-丙二醇的菌株。在研究的每种情况中甘油通过两步酶催化的系列反应被转化成 1,3-丙二醇。在第一步反应中脱水酶催化甘油转化为 3-羟基丙醛 (3-HP) 和水 (反应式 1)。在第二步反应中 3-HP 被 NAD<sup>+</sup> 相关的氧化还原酶还原成 1,3-丙二醇 (反应式 2)。



1,3-丙二醇不被进一步代谢,从而在培养基中高浓度富集。该整体反应消耗辅酶形式的还原型对等物,还原型  $\beta$ -烟酰胺腺嘌呤二核苷酸 (NADH), 其被氧化成烟酰胺腺嘌呤二核苷酸 (NAD<sup>+</sup>)。

从甘油生产 1, 3-丙二醇通常是在厌氧并且没有其他还原对等物接受体 (reducing equivalent acceptors) 的条件下用甘油作为唯一碳源得到的。在这些条件下, 在柠檬酸杆菌属, 梭菌属和克雷伯氏菌属中, 例如, 进行着一条平行的途径, 该途径首先由  $\text{NAD}^+$  (或  $\text{NADP}^+$ ) 相关的甘油脱氢酶将甘油氧化成二羟基丙酮 (DHA) (反应式 3)。DHA 被 DHA 激酶磷酸化成磷酸二羟基丙酮 (DHAP) 后 (反应式 4) 可被用于生物合成或通过, 例如, 糖酵解支持 ATP 的生成。



与 1, 3-丙二醇途径相反, 此途径可为细胞提供碳和能量并产生而不是消耗 NADH。

在肺炎克雷伯氏菌 (*Klebsiella pneumoniae*) 和弗氏梭菌 (*Citrobacter freundii*) 中编码功能上活性相关的甘油脱水酶 (dhaB), 1, 3-丙二醇氧化还原酶 (dhaT), 甘油脱氢酶 (dhaD) 和二羟基丙酮激酶 (dhaK) 的基因包含于 dha 调节子中。源自柠檬酸杆菌属和克雷伯氏菌属的 dha 调节子已经在大肠杆菌 (*Escherichia coli*) 中表达并且能将甘油转化为 1, 3-丙二醇。

1, 3-丙二醇的生物学方法生产需要甘油作为两步顺序反应的底物, 其中脱水酶 (通常为辅酶  $\text{B}_{12}$  依赖的脱水酶) 将甘油转化成中间产物 3-羟基丙醛, 然后 3-羟基丙醛被  $\text{NADH}^-$  (或  $\text{NADPH}$ ) 依赖的氧化还原酶还原。这些辅酶条件很复杂, 在使用了这种顺序反应生产 1, 3-丙二醇的工业方法中要求使用全细胞催化剂。一种从甘油生产 1, 3-丙二醇的含有辅酶  $\text{B}_{12}$  依赖的二元醇脱水酶的方法在 US 5, 633, 362 (Nagarajan et al.) 中有所描述。然而, 该方法不限于用甘油作为原料。葡萄糖和其他碳水化合物是适当的原料并且, 近来这些原料被用于 1, 3-丙二醇的生产。利用混合微生物碳水化合物被转化成 1, 3-丙二醇, 其中碳水化合物首先被一种微生物发酵成甘油, 然后被另一种微生物转化成 1, 3-丙二醇。US 5, 599, 689 (Haynie et al.)。为了简化以及经济起见能将碳水化合物转化成 1, 3-丙二醇的单一生物体是优选的。这样的生物体在 US 5, 686, 279 (Laffend et al.) 中有所描述。

一些细菌, 如沙门氏菌属或克雷伯氏菌属能合成辅酶  $\text{B}_{12}$  以使二元

醇或甘油脱水酶发挥作用，但是其他种属必须从细胞外转运  $B_{12}$ 。术语“ $B_{12}$ ”整体上用于指辅酶  $B_{12}$ ，上部轴向 5'-脱氧腺苷配基被其他配基取代（例如水-，氰-或甲基基团）的辅酶  $B_{12}$  衍生物；和基因类型，钴胺素（II）。

$B_{12}$  向细胞的转运存在两个问题。首先， $B_{12}$  分子太大了，不能通过外膜孔蛋白，从而需要特殊的外膜转运系统。第二，由于环境中的  $B_{12}$  很少，外外膜转运系统必须对  $B_{12}$  具有很高的亲和性并将其运送到质周腔以便随后被另一系统转运跨过内膜。对于不能合成  $B_{12}$  的咕啉环的大肠杆菌而言在一定条件  $B_{12}$  的外部供应对生长是必需的。这些要求可能是适度的；当存在功能性 MetH 基团时时需要大约 25 个  $B_{12}$  分子（甲基钴胺素），而对于依赖乙醇胺氨裂解酶的生长则需要大约 500 个辅酶  $B_{12}$  分子。

转运过程需要几种蛋白。66kDa 的外膜蛋白 BtuB 是腺苷-，水-，氰-和甲基钴胺素和相应的钴胺素（cobinamides）的高亲和性（ $K_d=0.3nM$ ）受体。在没有  $B_{12}$  或  $B_{12}$  水平很低（ $<1nM$ ）的条件下生长时每个细胞存在 ~200 拷贝的 BtuB。但是在含高水平（ $>0.1\mu M$ ） $B_{12}$  的培养基中生长时抑制了 BtuB 的合成，甚至在 5nM 的水平时摄入活力被抑制 80-90%。与沙门氏菌不同，大肠杆菌的 BtuB 不为需氧代谢所抑制。转运到质周腔需要 BtuB 和 26kDa 内膜蛋白 TonB 在能量依赖的途径中的相互作用，在该途径中还需要共转运钙。事实上， $B_{12}$  和 BtuB 的高亲和性结合是钙依赖的，有证据表明存在可逆的  $B_{12}$  依赖的钙结合位点，在 pH 6.6 和饱和水平的  $B_{12}$  的条件下对钙的  $K_d$  值为 ~30nM。上述对钙的亲和性随着 pH 的降低而降低。TonB 利用质子驱动力驱动转运所需要的结构变化。在没有 TonB 的情况下， $B_{12}$  穿过外膜的效率很低。TonB 还加强铁的外膜转运系统，包括 FepA 和 FhuA 系统。这样 BtuB 与这些系统竞争 TonB 的活性。没有蛋白合成时  $B_{12}$  转运的速率以 ~20 分钟的半衰期降低，这可归结于 TonB 活性的丧失。对  $B_{12}$  从 BtuB 向质周腔结合蛋白的转移的了解还很少，可能涉及 btuF 位点所编码的蛋白，至少在沙门氏菌中是这样的。

跨过内膜的转运是由 btuCED 操纵子所编码的 BtuC 和 BtuD 蛋白所介导的。BtuC 和 BtuD 类似于需要质周腔结合蛋白的转运蛋白，并且 BtuD 具有 ATP 结合位点。这两种基因的突变体的表型可通过适当升高

外界  $B_{12}$  的浓度而被校正，有人认为 BtuB/TonB 系统使得  $B_{12}$  在质周腔集中，从而促进  $B_{12}$  向细胞质的随机转运。BtuE 可能不参与转运，其功能未知。BtuCED 操纵子似乎为组成型表达，不为生长培养基中  $B_{12}$  的存在所调节。

上述转运途径可被总结为首先  $B_{12}$  与外膜蛋白 BtuB 结合，然后然后是与内膜蛋白 TonB 相互作用并进行能量依赖的移位，接着与质周腔的 BtuF 结合，最后转移到内膜蛋白 BtuCD 并移位到细胞质。

$B_{12}$  转运的一个重要控制机制是辅酶  $B_{12}$  对外膜蛋白 BtuB 水平的影响。由 btuR 基因所编码的 ATP: 类咕啉腺苷转移酶的活性导致了细胞辅酶  $B_{12}$  的形成。如上文所述，培养基中  $B_{12}$  的存在导致了 BtuB 功能的减弱，很有必要强调这种直接的抑制仅仅在辅酶  $B_{12}$  的情况中观察到而未在辅酶  $B_{12}$  前体中观察到，如通过向 btuR 缺陷菌株添加各种  $B_{12}$  分子所看到的那样。培养基中补充的辅酶  $B_{12}$  前体可能由于转变成辅酶  $B_{12}$  而导致抑制。调控似乎是通过改变信号的合成而不是信号的起始，因此使用外源启动子表达 btuB 并不一定能阻止辅酶  $B_{12}$  的调控。BtuB 调控不同寻常之处在于当 btuB 基因处于多拷贝质粒和单拷贝质粒上时阻遏似乎同样有效。这种表现上缺乏过量靶序列的滴定效应暗示着细胞中阻遏蛋白（辅酶  $B_{12}$ ）大大过量。

通过融合基因的研究发现转录水平和翻译水平的调控都适用于表达，将这些综合在一起，各种特征暗示了辅酶  $B_{12}$  和 mRNA 前导序列直接作用的机制。这种作用可能诱导 mRNA 折叠以稳定发夹结构从而阻止核糖体接近翻译的起始部位。BtuB 转录产物的相当一部分需要参与对其自身的表达和调控的控制，这意味着涉及前导序列和 btuB 编码区域的转录后事件影响了翻译的通读和起始。在氨基酸生物合成途径的衰减调控中已有报道转录区域参与调控，但是 btuB 调控的不同寻常之处在于重要的调节位点位于 btuB 的编码序列并且该调控同时影响转录和翻译。

本发明将要解决的问题是如何；利用一种含有辅酶  $B_{12}$  依赖的脱水酶的重组生物体通过生物学方法生产 1,3-丙二醇；上述重组生物体通过引入编码负责  $B_{12}$  转运的活性的基因而增强辅酶  $B_{12}$  的可获得性。

#### 发明概述

本专利的申请人通过提供能进行由脱水酶介导的从可发酵碳源向

1, 3-丙二醇直接转化的一种微生物而解决了上述问题, 其中通过引入  $B_{12}$  转运基因而增强了辅酶  $B_{12}$  对上述脱水酶的可获得性。在包括可发酵碳水化合物, 单碳底物和它们的混合物的一组较大的底物中优选的底物为葡萄糖和甘油。

本发明提供了一种 1, 3-丙二醇的生物生产方法, 包括: (i) 使转化的宿主细胞和至少一种可发酵碳源以及有效量的维生素  $B_{12}$  接触, 从而生产 1, 3-丙二醇, 上述转化的宿主细胞含有: (a) 至少一个拷贝编码具有脱水酶活性的蛋白的基因; (b) 至少一个拷贝编码具有氧化还原酶活性的蛋白的基因; (c) 至少一个拷贝编码维生素  $B_{12}$  受体前体蛋白的基因; (d) 至少一个拷贝编码维生素  $B_{12}$  转运系统通透酶蛋白的基因; (e) 至少一个拷贝编码具有维生素  $B_{12}$  转运 ATP-或 GTP 结合蛋白的基因; 其中(c), (d), 或(e)基因中的至少一个被引入到宿主细胞, 以及(ii)回收步骤(i)中生产的 1, 3-丙二醇。有效剂量的维生素  $B_{12}$  和存在的脱水酶的摩尔比为 0.1 至 10 倍。

本发明进而提供了表达脱水酶的转化宿主细胞, 含有(a)至少一个拷贝编码具有脱水酶活性的蛋白的基因; (b)至少一个拷贝编码具有氧化还原酶活性的蛋白的基因; (c)至少一个拷贝编码维生素  $B_{12}$  受体前体(B t u B)蛋白的基因; (d)至少一个拷贝编码维生素  $B_{12}$  转运系统通透酶蛋白的基因(B t u C); (e)至少一个拷贝编码具有维生素  $B_{12}$  转运 ATP-或 GTP 结合蛋白的基因(B t u D); 其中(c), (d), 或(e)基因中的至少一个被引入到宿主细胞。

#### 序列表简述

发明者按照专利申请中核苷酸和氨基酸序列标准表示法的规则(EPO局长决议附录 I 和 II, 发表于 OJ EPO, 12/1992 的增刊 2), 根据世界知识产权组织(WIPO)(1998)的 37 C. F. R. 1.821-1.825 和附录 A 和 B(对含有核苷酸和/或氨基酸序列的申请说明书的要求)以及 EPO 和 PCT 的序列表要求(Administrative Instruction 中的规则 5.2 和 49.5(a-bis) 和 Section 208 以及附件 C)提供了 25 个序列。根据在此引入作为参考的 Nucleic Acid Research 13:3021-3030(1985)和 Biochemical Journal 219(No. 2):345-373(1984)中所描述的 IUPAC-IYUB 的标准上述序列描述中含有编码核苷酸序列的单字母码和编码氨基酸序列的三字母码。

SEQ ID NO: 1 是大肠杆菌 butB 的核苷酸序列, 编码维生素 B<sub>12</sub> 受体前体蛋白。

SEQ ID NO: 2 是沙门氏菌 butB 的核苷酸序列, 编码维生素 B<sub>12</sub> 受体前体蛋白。

SEQ ID NO: 3 是大肠杆菌 butC 的核苷酸序列, 编码维生素 B<sub>12</sub> 转运系统通透酶蛋白。

SEQ ID NO: 4 是大肠杆菌 butD 的核苷酸序列, 编码维生素 B<sub>12</sub> 转运 ATP 结合蛋白。

SEQ ID NO: 5 是大肠杆菌 butE 的核苷酸序列, 编码维生素 B<sub>12</sub> 受体周质蛋白。

SEQ ID NO: 6 是 dha B1 的核苷酸序列, 编码甘油脱水酶 $\alpha$ 亚基。

SEQ ID NO: 7 是 dhaB2 的核苷酸序列, 编码甘油脱水酶 $\beta$ 亚基。

SEQ ID NO: 8 是 dhaB3 的核苷酸序列, 编码甘油脱水酶 $\gamma$ 亚基。

SEQ ID NO: 9 是 dhaT 的核苷酸序列, 编码克雷伯氏菌氧化还原酶。

SEQ ID NO: 10 是含有 dha 启动子的编码 PHK28-26 的一个 12.1kb EcoRI-SalI 片段的核苷酸序列。

SEQ ID NO: 11 是用于构建表达载体 pTacIQ 的多克隆位点和终止子的核苷酸序列。

SEQ ID NO: 12-23 是用于构建本发明载体的引物。

SEQ ID NO: 24 是 pCL1920 中的一段插入序列的核苷酸序列, 用于构建 dhaT 和 dha B(1, 2, 3) 的表达盒 (expression cassette)。

SEQ ID NO: 25 是链霉菌的葡萄糖异构酶启动子的核苷酸序列。

#### 发明详述

本发明提供了在单一重组生物生物体中利用生物学方法从可发酵碳源中生产 1, 3-丙二醇的方法。该方法利用了一种微生物, 该微生物含有编码甘油脱水酶, 1, 3-丙二醇氧化还原酶基因以及编码维生素 B<sub>12</sub> 受体前体 (BtuB) 的基因, 编码维生素 B<sub>12</sub> 转运系统通透酶蛋白的基因, 和编码维生素 B<sub>12</sub> 转运 ATP 结合蛋白 (BtuD) 的基因。上述微生物和碳底物接触, 1, 3-丙二醇分离自培养基。

本发明提供了一种快速, 便宜和环保的可用于生产聚酯和其他多聚体的 1, 3-丙二醇单体来源。

下列定义用于解释权利要求和说明书。

术语“维生素 B<sub>12</sub>受体”，“BtuB”或“外膜维生素 B<sub>12</sub>受体蛋白”指位于细菌外膜负责从培养基向周质空间转运辅酶 B<sub>12</sub>，氰钴氨素，水合钴氨素，甲基钴氨素和钴啉醇酰胺的多肽。为了本发明的目的，BtuB 包括，例如，由大肠杆菌的 btuB 基因 (GenBank M10112) (SEQ ID NO:3) 所编码的蛋白以及鼠伤寒沙门氏菌 (*Salmonella typhimurium*) (GenBank M89481) (SEQ ID NO:2)。

术语“BtuC”或“维生素 B<sub>12</sub>转运系统通透酶蛋白”指位于细菌内膜的多肽，和 BtuD 一起从周质空间向细胞质转运转维生素 B<sub>12</sub>和辅酶 B<sub>12</sub>。BtuC 包括，例如，由大肠杆菌的 btuC 基因 (GenBank M14031) (SEQ ID NO:3) 编码的多肽。

术语“BtuD”或“维生素 B<sub>12</sub>转运 ATP 结合蛋白”指位于细菌内膜的多肽，和 BtuC 一起从周质空间向细胞质转运转维生素 B<sub>12</sub>和辅酶 B<sub>12</sub>。BtuD 包括，例如，由大肠杆菌的 btuD 基因 (GenBank M14031) (SEQ ID NO:4) 编码的多肽。

术语“BtuE”指位由大肠杆菌的 btuE 基因 (GenBank M14031) (SEQ ID NO:5) 编码的多肽，是上述转运系统的辅助成分。

术语“甘油脱水酶”或“脱水酶”指负责辅酶 B<sub>12</sub> 依赖的酶活性的多肽，上述辅酶 B<sub>12</sub> 依赖的酶活性能将甘油异构化或转化产物成为 3-羟基丙酮。为了本发明的目的，脱水酶包括优选底物分别为甘油和 1, 3-丙二醇的甘油脱水酶 (GenBank U09771, U30903) 和二醇脱水酶 (Genbank D45071)。肺炎克雷伯氏菌 ATCC 25955 的甘油脱水酶由基因 dhaB1, dhaB2 和 dhaB3 编码，分别用 SEQ ID NO: 6, 7 和 8 标识。基因 dhaB1, dhaB2 和 dhaB3 分别编码甘油脱水酶的  $\alpha$ ,  $\beta$  和  $\gamma$  亚基。甘油脱水酶和二醇脱水酶是复合体 (亚基组成为  $\alpha_2\beta_2\gamma_2$ )，和辅酶 B<sub>12</sub> 以 1: 1 的比例结合。

“有效剂量”的辅酶 B<sub>12</sub> 前体 (或维生素 B<sub>12</sub>) 的意思为系统中存在的辅酶 B<sub>12</sub> 前体 (或维生素 B<sub>12</sub>) 相对脱水酶的摩尔比为 0.1 至 10。

术语“氧化还原酶”或“1, 3-丙二醇氧化还原酶”指负责能催化 3-羟基丙醛还原成 1, 3-丙二醇的活性的多肽。1, 3-丙二醇氧化还原酶包括，例如，由 dhaT 基因编码的多肽被标识为 SEQ ID NO: 9。

术语“辅酶 B<sub>12</sub>”和“腺苷基钴氨素”和 5'-脱氧腺苷基钴氨素可以互换使用。羟基钴氨素是辅酶 B<sub>12</sub> 的衍生物，其中上轴向的 5'-脱氧腺苷基配基为羟基所取代。水合钴氨素是羟基钴氨素的质子化形式。甲基钴氨素是辅酶 B<sub>12</sub> 的衍生物，其中上轴向的 5'-脱氧腺苷基配基为甲基所取代。术语“氰钴氨素”用于指辅酶 B<sub>12</sub> 的上轴向 5'-脱氧 5'-腺苷基配基为氰所取代的衍生物。术语“维生素 B<sub>12</sub>”和“B<sub>12</sub>”用于可互换地整体上指辅酶 B<sub>12</sub>；辅酶 B<sub>12</sub> 的上轴向 5'-脱氧腺苷基配基为其他配基所取代的衍生物，例如，水合-，氰-，或甲基；以及基因类型，钴胺素（II）。术语“辅酶 B<sub>12</sub> 前体”指辅酶 B<sub>12</sub> 的上轴向 5'-脱氧腺苷基配基被取代的衍生物。“有效剂量”的辅酶 B<sub>12</sub> 前体的意思为系统中存在的辅酶 B<sub>12</sub> 前体（或维生素 B<sub>12</sub>）相对脱水酶的摩尔比约为 0.1 至 10。

术语“多肽”和“蛋白”可以互换使用。

术语“可发酵碳底物”和“可发酵碳源”指能被本发明的宿主生物体代谢的碳源，尤其是选自单糖，寡聚糖，多糖，甘油，二羟基丙酮和单碳底物以及它们混合物的碳源。

术语“宿主细胞”或“宿主生物体”指能接受外源或异源基因或多拷贝内源基因并表达这些基因产生活性基因产物的微生物。

“基因”指表达特定蛋白的核酸片段，包括编码序列之前（5'非编码序列）和之后（3'非编码序列）的调节序列。“天然基因”指在自然中发现的带有自身调节序列的基因。“融合基因”指任何不是天然基因的，含有不是在自然中发现的调节和编码序列的基因。因此，任何基因可以含有源自不同来源的调节序列和编码序列，或调节序列和编码序列源自同一来源但是排列的方式和自然中发现的方式不同。“内源基因”指在一个生物体基因组的天然位置的天然基因。“外源”或“异源”基因指通常不会在宿主生物体中发现的基因，但是通过基因转化被引入宿主生物体。外源基因可以含有插入到非天然有机体，或融合基因，的天然基因。“转基因”是通过转化的方法被引入到基因组的基因。

术语“编码”（encoding and coding）指一个基因通过转录和翻译的机制产生蛋白质的过程。编码特定氨基酸序列的过程包括 DNA 序列，DNA 序列可能发生不会导致所编码的氨基酸变化的碱基变化，

或发生可能发生改变一个或几个编码的氨基酸但是不影响 DNA 序列所编码的蛋白质的功能的碱基变化。因此认为本发明不仅仅包括特定的示例性序列。对序列的修改，如在序列中进行基本上不影响所得到的蛋白分子的功能的产生静息变化的缺失、插入、或替代也是本发明所预期的。例如，基因序列中反应了遗传密码的简并性的改变或在特定位点导致产生化学对等氨基酸的变化是预期的。因此，一个丙氨酸，一种疏水氨基酸，的密码子可以为编码另一疏水性较弱的残基（如甘氨酸）的密码子或为疏水性更强的残基（如缬氨酸，亮氨酸或异亮氨酸）的密码子所取代。类似地，导致一个带负电残基取代另一带负电残基（例如天冬氨酸取代谷氨酸）的变化，或一个带正电残基取代另一带正电残基（例如赖氨酸取代精氨酸）的变化也可预期产生生物学上对等的产物。可以预期导致蛋白 N 末端或 C 末端部分变化的核苷酸变化可能改变蛋白的活性。在一些情况下制备序列的突变体以研究变化对蛋白生物学活性的影响也是合乎需要的。以上列出的这些变化都在本领域的常规技术之内，确定在编码的产物中保留生物学活性也是。进而，本领域的技术人员认识到本发明所包括的序列也可被定义为能与此处列出的示例性序列在严紧条件下（ $0.1\times\text{SSC}$ ， $0.1\%\text{SDS}$ ， $65^\circ\text{C}$ ）杂交的序列。

术语“表达”指从编码基因产物的基因转录和翻译得到基因产物。

术语“质粒”“载体”和“盒（cassette）”指通常带有基因的另加染色体元件，上述基因不是细胞核心代谢的一个部分，并通常为环状双链 DNA 分子的形式。这样的元件可以是自动复制的序列，基因组整合序列，噬菌体或核苷酸序列，线性或环状的，单链或双链的 DNA，源自任何来源，其中许多核苷酸序列被连接或重组成独特的构建体，该构建体能将编码选择的基因产物的 DNA 序列和启动子片段引入细胞。“转化盒”指含有外源基因以及外源基因之外的元件，能促进特定宿主细胞的转化的特定载体。“表达盒”指含有外源基因和含有外源基因以及外源基因之外的元件并允许增强该基因在其宿主中表达的特定载体。

术语“转化”和“转染”指吸收核酸后细胞获得新基因。获得的基因可以被整合入染色体 DNA 或作为染色体外复制序列被引入。术语“转化体”指转化的产物。

术语“遗传上发生改变的”指通过转化或突变改变遗传物质。

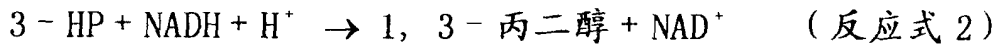
本发明涉及构建一种生产生物体，其含有将可发酵碳底物转化为 1, 3-丙二醇所必需的遗传机制，以及编码细胞内转运维生素 B<sub>12</sub> 所必需的酶的基因。参与 1, 3-丙二醇生产的基因包括脱水酶基因（通常为甘油或二元醇脱水酶）和氧化还原酶基因以及其它预期能辅助脱水酶的组装或维持其稳定性的蛋白的基因。这些基因可以是转基因并被引入到宿主细胞，或可以是内源基因。负责细胞内维生素 B<sub>12</sub> 转运的基因包括至少一种编码维生素 B<sub>12</sub> 受体前体蛋白的基因（BtuB），至少一种编码维生素 B<sub>12</sub> 转运系统通透酶的基因（BtuC）以及至少一种编码维生素 B<sub>12</sub> 转运 ATP 结合蛋白的基因（BtuD）。这些基因中至少一个是转基因并被引入生产细胞。然后在适当的条件下生长转化后的生产细胞以生产 1, 3-丙二醇。

含有编码将碳源底物转化成 1, 3-丙二醇的酶学途径所必需的基因的重组生物体的构建可利用本领域中众所周知的技术。在本发明中分离自天然宿主如克雷伯氏菌属的编码甘油脱水酶(dhaB)和 1, 3-丙二醇氧化还原酶(dhaT)的基因, 以及分离自天然宿主如大肠杆菌和鼠伤寒沙门氏菌的编码 BtuB(btuB), BtuC(btuC), BtuD(btuD) 和 BtuE(btuE)α或 FM5; 肺炎克雷伯氏菌菌株 ATCC25955; 产酸克雷伯氏菌(K. oxytoca)菌株 ATCC8724 或 M5a1, 啤酒糖酵母(S. cerevisiae)菌株 YPH499, 巴斯德毕赤酵母(P. pastoris)菌株 GTS115, 或黑曲霉(A. niger)菌株 FS1.

DhaB, dhaT 的基本原理

从葡萄糖生产 1, 3-丙二醇可通过以下一系列步骤来完成。这一系列代表了许多本领域中技术熟练人员所已知的途径。葡萄糖被糖酵解途径的酶通过许多步骤转化成磷酸二羟基丙酮(DHAP)和 3-磷酸甘油醛(3-PG)。DHAP 水解成二羟基丙酮 DHA 后还原形成甘油, 或者 DHAP 还原成甘油 3-磷酸(G3P)后水解形成甘油。水解步骤可为任何细胞磷酸酶催化, 已知它们对于它们的底物是非特异性的, 或者该活性可通过重组被引入宿主。还原步骤可以由 NAD<sup>+</sup> (或 NADP<sup>+</sup>) 相关的宿主酶催化或该活性可通过重组被引入。值得注意的是 dha 调节子含有催化反应式 3 的可逆反应的脱氢酶(E. C. 1. 1. 1. 6):





上文详细描述了甘油通过中间产物 3-羟基丙酮(3-HP)被转化成 1,3-丙二醇。中间产物 3-HP 是从甘油, 反应式 1, 通过脱水酶得到的, 上述脱水酶为宿主所编码或可通过重组被引入宿主。该脱水酶可以是甘油脱水酶(E. C. 4.2.1.30) 二元醇脱水酶(E. C. 4.2.1.28) 或任何其他能催化这种转变的酶。甘油脱水酶, 而不是二醇脱水酶, 是由 dha 调节子所编码的。1,3-丙二醇是由  $\text{NAD}^+$  (或  $\text{NADP}^+$ ) 相关的宿主酶从 3-HP 产生, 反应式 2, 上述宿主酶的活性也可通过重组被引入宿主。产生 1,3-丙二醇的终反应可由 1,3-丙二醇脱氢酶(E. C. 1.1.1.202) 或其它醇脱氢酶催化。

dha 调节子由几种功能元件组成包括编码二羟基丙酮激酶的 dhaK, 编码甘油脱氢酶的 dhaD 和编码调节蛋白的 dhaR, 编码 1,3-丙二醇氧化还原酶的 dhaT 以及分别编码该酶  $\alpha$ 、 $\beta$ 、 $\gamma$  亚基的 dhaB1, dhaB2 和 dhaB3。基因产物被命名为蛋白 X, 蛋白 1, 蛋白 2 和蛋白 3 (分别相应于 dhaBX, orfY, orfX, 和 orfW) 也被编码于 dha 调节子内。这些基因产物的确切功能还不是很清楚, 这些基因是与甘油脱水酶(dhaB) 或 1,3-丙二醇氧化还原酶(dhaT) 相关的, 并且已知对于 1,3-丙二醇的生产是有用的。结合于甘油脱水酶的辅酶  $\text{B}_{12}$  有时发生不可逆的裂解形成与脱水酶紧密结合的无活性的修饰辅酶。该酶的重新活化是通过结合的修饰的辅酶与自由的完整的辅酶  $\text{B}_{12}$  的交换。蛋白 X 和其它蛋白 1, 蛋白 2 和蛋白 3 中的至少一种参与这种交换过程。(参见 USSN 08/969,683)。在独立的二醇脱水酶系统中被命名为 ddrA 和 ddrB 的基因, 相应于编码蛋白 X 和蛋白 2 的基因, 被认为参与了该交换过程。Morri et al., J. Biol. Chem. 272, 32034-32041(1997)。

有人认为甘油-3-磷酸脱氢酶和甘油-3-磷酸酶在葡萄糖转化为甘油中特别有效, 在 1,3-丙二醇的生产中是必需的。术语“甘油-3-磷酸脱氢酶”指负责催化磷酸二羟基丙酮(DHAP) 向甘油-3-磷酸(G3P) 转化的酶活性的多肽。在体内 G3PDH 可以是 NADH、NADPH 或 FAD 依赖的。NADH 依赖的酶(EC. 1.1.1.8) 是由, 例如几种基因编码的, 包括 GPD1 (GenBank Z74971x2) 或 GPD2 (GenBank Z35169x1)

或 GPD3(GenBank G984182), 或 DAR1 (GenBank Z74071x2)。NADPH 依赖的酶(EC 1.1.1.94)是由 gpsA(Genbank U321643, (cds 197911-196892)和 G4667646 和 L45246)。FAD 依赖的酶 (EC1.1.99.5) 是由 GUT2 ( GenBank Z47047x23 ) 或 glpD(GenBank G147838) 或 glpABC(GenBank M20938) 编码。术语“甘油-3-磷酸酶”指负责催化甘油-3-磷酸向水和甘油和无机磷酸转化的酶活性的多肽。甘油-3-磷酸酶是由, 例如, GPP1 (GenBank Z47041x125), 或 GPP2(GenBank U18813x11) 编码。

### 基因分离

从细菌基因组获得所需基因的方法是常规的并且在分子生物学领域是已知的。例如, 如果该基因的序列是已知的, 则可以通过限制性内切酶消化建立基因组文库并用与所需序列互补的探针筛选。一旦分离得到了该序列, 则可以通过常规的引物定位的扩增方法如聚合酶链式反应(PCR) (US 4, 683, 202) 扩增 DNA 以得到适用于用恰当的载体转化的一定量的 DNA。

替代地, 可以建立粘粒(cosmid)文库, 在粘粒文库中大片段的基因组 DNA 可被装入载体并用于转化使得的宿主。粘粒载体的独特之处在于能容纳大量 DNA。通常, 粘粒载体具有至少一个可拷贝的 cos DNA 序列, 该序列对于外源 DNA 的装入和随后的环化是必需的。除了 cos 序列外这些载体还含有复制起点如 ColE1 和药物抗性标记如抗氨苄青霉素或新霉素的基因。用粘粒载体转化适当的细菌宿主的方法在 Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition(1989) Cold Spring Harbor Laboratory Press, Cold Spring Hrbor, NY(1989)中有很好的描述。

通常对于克隆的粘粒而言通过适当的限制性内切酶外源 DNA 被分离并连接到和粘粒载体的 cos 区域相邻的区域上。然后含有线性外源 DNA 的粘粒载体与 DNA 包装载体如噬菌体 $\lambda$ 反应。在包装过程中 cos 位点被切开并且外源 DNA 被包装成细菌病毒颗粒的头部。这些颗粒然后被用于转染适当的宿主菌如大肠杆菌一旦注入细胞, 外源 DNA 在 cos 粘性末端的作用下环化。通过这种方式大片段的外源 DNA 可被引入重组宿主细胞并在其中表达。

编码甘油脱水酶(dhaB)基因以及 1, 3-丙二醇氧化还原酶基因(dhaT)

## 的分离和克隆

dhaB 和 dhaT 的鉴定和分离的方法基本上按照 US 5,686,276 中所描述的进行, US 5,686,276 在此引入作为参考。在本发明的情况下, 粘粒载体和粘粒转化方法被用于从已知具有能将甘油加工为 1,3-丙二醇的基因的细菌种属中克隆大片段基因组 DNA。分析了两个 1,3-丙二醇阳性转化子并且 DNA 测序表明和弗氏柠檬酸杆菌中的甘油脱水酶基因 (dhaB) 具有广泛的同源性, 这表明这些转化子含有表明甘油脱水酶基因的 DNA。dhaB 和 dhaT 被分离并克隆到适当的表达盒中以在具有编码 B12 转运功能的基因的重组宿主中共表达。

虽然本发明直接使用了分离自克雷伯氏菌粘粒的基因, 脱水酶基因的替代来源还包括, 但不限于, 柠檬酸杆菌属, 梭菌属和沙门氏菌。  
B12 转运基因的基本原理

腺苷基-钴胺素 (辅酶 B12) 是甘油脱水酶活性的必需辅酶。该辅酶是已知的最复杂的非多聚体天然产物, 其合成直接使用了约 30 个基因的产物。辅酶 B12 的合成发现于原核生物, 其中一些能从头合成该复合物, 而另一些能进行部分反应。例如, 大肠杆菌不能制造咕啉环结构, 但能催化咕啉醇酰胺向类咕啉的转化并能引入 5'-脱氧腺苷基。

B12 向大肠杆菌内的转运可能是产生具有功能的 Dha B 酶的限制性因素, 在这样的情况下提高细胞内辅酶 B12 的可获得性对于优化甘油脱水酶活性 (并且, 最终地, 1,3-丙二醇的生产) 是必需的。这可通过提高 B12 向细胞内的转运速率而实现。考虑到辅酶 B12 是 btuB 表达的抑制剂, 并且考虑到发酵时辅酶 B12 所需要的水平, 很可能由于其周转或者细胞分裂造成的 BtuB 的稀释 B12 转运随时间降低。细胞中可获得的辅酶 B12 库将受到吸收速率或 BtuB mRNA 的影响以及需要辅酶 B12 的 DhaB 和 mRNA 以及 DhaB 浓度的影响。由于使用富含 B12 的培养基时吸收减少, 决定吸收机制是否恢复的一个重要因素是辅酶 B<sub>12</sub> 在其对 btuB 的调节作用以及 DhaB 之间的分配。这表现一个不同寻常的问题: 所需要的辅酶 (辅酶 B<sub>12</sub>) 对其自身进行限制。使用辅酶 B<sub>12</sub> 前体取代辅酶 B<sub>12</sub> 将减轻此问题, 但这只有暂时的效果, 因为被转运进去的前体将被 btu-R 编码的腺苷基转移酶转化为辅酶 B<sub>12</sub>。一种避免该基因调节问题的方法是使 BtuB 的合成与辅酶 B<sub>12</sub> 的调节解偶联。通过

在多拷贝质粒上克隆而扩增 *btuB* 的表达将导致  $B_{12}$  与膜的结合增加并提高吸收的速率，并且如果使用了 *btuB* 天然启动子也将使 *BtuB* 的合成与辅酶  $B_{12}$  的调节解偶联。

$B_{12}$  向细胞的转运需要特定的转运系统。该转运过程需要几种蛋白。66kDa 的外膜蛋白 *BtuB* 是腺苷基-，水合-，氟-和甲基-钴氨素或相应的钴啉醇酰胺类的受体。转运入质周腔需要 *BtuB* 与 26kDa 内膜蛋白 *TonB* 通过能量依赖的过程相互作用。跨内膜的转运是由 *btuCED* 操纵子编码的 *BtuC* 和 *BtuD* 蛋白介导。*BtuC* 和 *BtuD* 类似于需要质周腔结合的蛋白的转运蛋白，并且 *BtuD* 具有一个 ATP 结合位点。转运途径可被归结成首先  $B_{12}$  与外膜蛋白 *BtuB* 结合，然后与内膜蛋白 *TonB* 相互作用，然后是能量依赖的转运以及与质周腔 *Btuf*（鼠伤寒沙门氏菌中）的结合，最后转移到内膜蛋白 *BtuCD* 并转运到细胞质中。通过多拷贝质粒上的克隆扩增 *btuCED* 的表达导致了  $B_{12}$  与膜的结合增加并增加了其被吸收入细胞的速率。

#### $B_{12}$ 转运基因的分离和表达

为 4 种  $B_{12}$  转运基因，*btuB*，*btuC*，*btuD* 和 *btuE* 构建了能在大肠杆菌中作为复制元件存在的表达质粒。利用对基因特异的引物和大肠杆菌染色体 DNA 分离了 4 种基因。这 4 中基因在表达质粒上被组装在一起。所有的表达质粒都用 *trc* 启动子进行转录，用天然 *btu* 核糖体结合位点进行翻译。每个质粒还含有 1) 用于在含氨基青霉素的培养基中进行选择的  $\beta$ -内酰胺酶基因或 2) 用于在含氯霉素的培养基中进行选择的氯霉素乙酰转移酶基因之中的一个。质粒的复制起点是 *ColE1* 或 *p15A*。

#### 宿主细胞

用于通过编码脱水酶的基因以及负责细胞内  $B_{12}$  转运的基因的共表达而重组生产 1, 3-丙二醇的适当的宿主细胞可以是原核细胞或真核细胞，限制条件仅仅是它们表达活性酶的能力。优选的宿主将是这些尤其有用于生产 1, 3-丙二醇或甘油的宿主，例如柠檬酸杆菌属，肠杆菌属，梭菌属，克雷伯氏菌属，气杆菌属 (*Aerobacter*)，乳杆菌属，曲霉属 (*Aspergillus*)，糖酵母属 (*Sacharomyces*)，裂殖糖酵母属 (*Schizosaccharomyces*)，结合糖酵母属 (*Zygosaccharomyces*)，毕赤酵母属 (*Pichia*)，克鲁维氏酵母属 (*Kluyveromyces*)，假丝酵母属

(Candida), 汉逊氏酵母属 (Hansenula), 德巴利氏酵母 (Debaryomyces), 毛霉属 (Mucor), 球拟酵母属 (Torulopsis), 甲基杆菌 (Methylobacter), 大肠杆菌属, 沙门氏菌属 (Salmonella), 芽孢杆菌属 (Bacillus), 链霉菌属 (Streptomyces) 和假单胞菌类 (Pseudomonas)。

大肠杆菌, 糖酵母类和克雷伯氏酵母类是尤其优选的宿主。已知肺炎克雷伯氏菌的菌株在用甘油作为唯一碳源时能生产 1, 3-丙二醇。预期克雷伯氏菌属可通过基因手段被改变成能从单糖, 寡糖, 多糖或单碳底物生产 1, 3-丙二醇。

#### 载体和表达盒

本发明提供了许多适用于编码适当的脱水酶的基因和执行 B<sub>12</sub> 向适当的宿主细胞转运功能的基因的克隆转化和表达的载体、转化盒和表达盒。适用的载体将与所使用的细菌相兼容。适当的载体可源自, 例如, 细菌、病毒 (例如 T7 噬菌体或源自 M13 的噬菌体) 粘粒、酵母或植物。使用和获得这样的载体的方法在本领域中是已知的 (Sambrook et al., Molecular Cloning: A Laboratory Manual- volumes 1, 2, 3 (Cold Spring Laboratory, Cold Spring Harbor, NY, (1989))。

通常载体或盒含有引导相关基因的转录和翻译的序列、选择标记和允许自主复制或染色体整合的序列。适合的载体基因的 5'端区域带有对转录启动的控制, DNA 片段的 3'端控制转录的终止。最为优选的是两种控制区域都源自和转化宿主细胞同源的基因, 尽管认为这样的控制区域不需要源自被选择作为生产宿主的特定种的天然基因。

可用于启动本发明的相关基因在所需的宿主细胞中表达的启动控制区域或启动子有很多, 并且为本领域的技术熟练人员所熟悉。基本上能驱动这些基因表达的任何启动子都适用于本发明, 包括但不限于, CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI (可用于在糖酵母属中表达); AOX1 (可用于在毕赤酵母中表达); 和 lac, trp,  $\lambda P_R$ , T7, tac, 和 trc (可用于在大肠杆菌中表达)。

终止控制区域也可源自各种对优选的宿主天然的基因。供选地, 终止位点可以是不必要的; 然而, 如果包含的话是最为优选的。

为了酶的有效表达, 编码该酶的 DNA 通过起始密码子可操作地连

接到表达控制区域上以使得表达导致形成适当的信使 RNA。

适当宿主的转化和基因的表达以生产 1, 3-丙二醇

一旦构建了适当的盒，它们即被用于转化适当的宿主细胞。可通过已知的方法，例如通过转化（例如用钙渗透细胞，电穿孔）或通过用重组噬菌体病毒转染（Sambrook et al., supra）将分别含有或同时含有负责细胞内 B<sub>12</sub> 转运的基因以及负责甘油脱水酶（dhaB），和 1, 3-丙二醇氧化还原酶基因（dhaT）的上述盒导入宿主细胞。

在本发明中，含有编码甘油脱水酶（dhaB），1, 3-丙二醇氧化还原酶（dhaT），BtuB（btuB），BtuC（btuC），BtuD（btuD）和 BtuE（btuE）的 *E. coli* FM5 被用于从培养基向细胞质转运输维生素 B<sub>12</sub> 或辅酶 B<sub>12</sub> 以使得甘油脱水酶发挥功能。

培养基和碳源底物

本发明中发酵培养基必需含有适当的碳源底物。适当的碳源底物包括单不限于甘油、二羟基丙酮、单糖如葡萄糖和果糖，寡糖如乳糖或蔗糖，多糖如淀粉或纤维素，或它们的混合物，以及源自补充给料的未纯化混合物如奶酪乳清渗透物（cheese whey permeate），谷物浸泡液，甜菜糖蜜和大麦麦芽。此外，碳源底物也可以是向关键生化中间产物的转化已经被证明的单碳底物（如二氧化碳或甲醇）。

已有报道在甲基营养酵母（Methylotrophic yeasts）（Yamada et al., *Agric. Biol Chem.*, 53(2) 541-543(1989)）和细菌中（Hunter et al, *Biochemistry*, 24, 4148-4155）从单碳源生产甘油。这些生物体能同化氧化状态在甲烷和甲酸之间的单碳化合物并产生甘油。碳同化的途径可以通过核酮糖单磷酸途径，通过丝氨酸途径或通过木酮糖单磷酸途径（Gottschalk, *Bacterial Metabolism*, Second Edition, Springer-Verlag: New York(1986)）。核酮糖单磷酸途径需要缩合甲酸和核酮糖-5-磷酸形成 6 碳糖，6 碳糖转化为果糖并最终成为 3 碳产物甘油醛-3-磷酸。类似地，丝氨酸途径将单碳化合物通过亚甲基四羟基叶酸同化进入糖酵解途径。

除了利用单碳或二碳底物，已知甲基营养生物体还利用许多其它含碳化合物如甲胺，葡萄糖胺和许多氨基酸进行代谢活动。例如，已知甲基营养酵母利用甲胺的碳形成海藻糖或甘油（Bellion et al., *Micro. Growth Cl Compd.*, [Int. Symp.] 7<sup>th</sup>(1993), 415-32.

Editor(s): Murrell, J. Collin; Kelley, Don p. Publisher: Intercept, Andover, UK)。类似地, 几种假丝酵母的种能代谢丙氨酸或油酸 (Sulter et al., Arch. Microbiol., 153(5), 485-9(1990))。因此, 本发明中所利用的碳的来源可以包括许多含碳底物并且仅仅受宿主组织的需要的限制。

尽管认为所有上述的碳源底物及它们的混合物都适用于本发明, 优选的底物是甘油, 二羟基丙酮, 单糖, 寡糖, 多糖, 和单碳底物。更为优选的是如葡萄糖、果糖、蔗糖这样的糖类和单碳底物如甲醇和二氧化碳。最为优选的是蔗糖。

除了适当的碳源外, 发酵培养基还必须含有适当的矿物质, 盐, 辅酶, 缓冲剂和适于培养物的生长以及促进甘油产生所需要的酶途径的其它组分, 这些是本领域的技术熟练人员所已知的。尤其值得注意的是 Co(II) 盐和辅酶 B<sub>12</sub> 前体。例如, 大肠杆菌和真核细胞不能从头合成辅酶 B<sub>12</sub>, 但能利用辅酶 B<sub>12</sub> 前体。优选的辅酶 B<sub>12</sub> 前体是氰钴氨素和羟基钴氨素。宿主细胞内辅酶 B<sub>12</sub> 的量在摩尔浓度上大约等于脱水酶的量是赫赫需要的。

#### 培养条件

通常, 细胞在适当的培养基中在 30℃ 生长。本发明中优选的培养基通常为商业上已经制备好的培养基如 Luria Bertani(LB) 肉汤培养基, Sabouraud Dextrose(SD) 肉汤培养基或酵母麦芽抽提物 (YM) 肉汤培养基。也可使用其它描述的或合成的培养基, 特定微生物生长的适当培养基对于在微生物或发酵科学中的技术熟练人员是已知的。已知的直接或间接调节分解代谢抑制作用的制剂, 例如, 环腺苷 3':5'-单磷酸, 也可被加入到反应培养基中。类似地, 也可以联合使用已知调节酶活性增强甘油的产生的制剂 (例如, 亚硫酸盐, 亚硫酸氢盐和碱) 以及基因操作, 或作为基因操作的替代手段。

适当的发酵 pH 范围是在 pH5.0 至 pH9.0 之间, 而 pH6.0 至 pH8.0 是初始条件的优选范围。

反应可以在需氧或厌氧的条件下进行, 其中厌氧或微厌氧的条件是优选的。

#### 发酵

本发明的实践可以用批量方式, 进料批量方式 (Fed-Batch), 或

连续过程发方式，或任何适当的已知发酵方式进行。进而，认为细胞可被固定到底物上作为全细胞催化剂并在发酵条件下进行 1, 3-丙二醇的生产。

此处示例的本方法为发酵的批量方法。一种经典的批量方法是密闭的系统，其中培养基的成分在发酵起始时是固定的并且在发酵过程中不进行人工改变。这样，在发酵的开始时用所需的一种或几种生物体接种培养基并且允许发酵进行，并且不向系统添加任何东西。然而，通常批量发酵的“批量”是相对于碳源的添加而言的，并且经常尝试调节控制因子如 pH 和氧浓度。批量系统中的代谢物和生物数量组分经常改变，直至发酵停止。在批量系统中培养细胞经过静态延迟期到高生长的对数生长期并最终到静止期，在静止期中生长速度减慢或停止。如果不进行处理，静止期的细胞将最终死亡。对数生长期的细胞通常负责生产终产物或中间产物。

标准批量系统的一种变化是进料-批量发酵系统，该系统也适用于本发明。在这种经典批量系统的变型中底物随着发酵过程的发展而添加。当代谢产物抑制易于抑制细胞的代谢以及需要限制培养基中底物的数量时进料-批量系统是有用的。测量进料-批量系统中真正的底物浓度是困难的，因此是在可测量因素如 pH，溶解氧和废气如 CO<sub>2</sub> 的分压的基础上经常估计的。批量和进料批量发酵在本领域中是常用并且众所周知的，例子可参见 Brock. 上文。

该方法也适用于连续发酵方法。连续发酵是开发系统其中给定的发酵培养基被连续添加到生物反应器中并同时取出等量培养基进行处理。连续发酵通常维持着高密度的培养物，其中细胞主要处于对数期生长。

连续发酵允许调节影响细胞生长或终产物浓度的一个因素或任何数量的因素。例如，一种方法将按照固定的水平维持有限的营养如碳源或氮水平并使所有其它参数处于适中的水平。在其它系统中影响生长的因素的数量可以被连续地改变而通过培养基混浊度测量得到的细胞浓度保持恒定。连续系统努力维持恒定生长状态，因此由于培养基的取出而导致的细胞损失必须和发酵中细胞的生长速率保持平衡。连续发酵方法中调节营养和生长因素的方法以及使产物形成速率最大化的技术在工业微生物的领域中是广为人知的。许多方法在 Brock, 上

文中有详细描述。

### 1, 3-丙二醇的鉴定和纯化

从发酵培养基中纯化 1, 3-丙二醇的方法是已知的。例如, 例如通过将反应混合物用有机溶剂进行抽提, 蒸馏和柱层析 (US 5, 356, 812)。对于此方法特别好的有机溶剂是环己烷 (US 5, 008, 473)。

可以通过将培养基直接进行高压液相色谱 (HPLC) 分析以鉴定 1, 3-丙二醇。在本发明的一种优选的方法中发酵培养基在分析型离子交换柱上用 0.01N 硫酸作为流动相以等度洗脱的方式进行。

本发明将在随后的实施例中进一步定义。应当理解这些实施例在说明本发明的优选实施方案的同时仅仅是以举例说明的方式给出的。从以上的讨论和这些实施例, 本领域中技术熟练的人员将确认本发明的本质特征, 并且能在不偏离本发明的精神和范围的前提下进行本发明各种变化和修正以使之适用于各种用途和条件。

## 实施例

### 一般方法

磷酸化、连接和转化的方法在本领域中是众所周知的。适用于以下实施例的技术可在 Sambrook, J. et al., *Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY(1989)中找到。

适用于维持和生长细菌培养物的材料和方法在本领域中是众所周知的。适用于以下实施例的技术可在 *Manual of Methods for General Bacteriology* (Philipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994) 或 Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology, Second Edition*, Sinauer Associate, Inc., Sunderland, MA (1989) 中找到。除非特别说明, 所有用于细菌细胞的维持和生长的试剂和材料都可从 Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), 或 Sigma Chemical Company (St. Louis, MO) 得到。

缩写的含义如下: “h”的意思是小时, “min”的意思是分钟, “sec”

的意思是秒，“d”的意思是天，“mL”的意思是毫升，“L”的意思是升。

#### 1, 3-丙二醇的分离和鉴定

甘油向 1, 3-丙二醇的转化是通过 HPLC 监测的。进行分析所用的常规技术和材料对于层析领域中的技术熟练人员都是可以得到的。一种适用的方法使用了 Water Maxima 820 HPLC system 用 UV (210nm) 和 RI 进行检测。样品被注射入配备了 Shodex SH-1011P precolumn (6mm×50mm) 的 Shodex SH-1011 层析柱 (8mm×300mm, 购自 Waters, Milford, MA) 中, 温度控制在 50℃, 用 0.01N H<sub>2</sub>SO<sub>4</sub> 作为流动相, 流速 0.5mL/min。当需要定量分析时以三甲基作为外标。通常, 甘油 (RI 检测), 1, 3-丙二醇 (RI 检测) 和三甲基乙酸 (UV 和 RI 检测) 的停留时间分别为 20.67min, 26.08min 和 35.03min。

1, 3-丙二醇的生产通过 GC/MS 确认。分析用的常规技术和材料对于 GC/MS 中的技术熟练人员都是可以得到的。一种适用的方法使用了连接到 Hewlett Packard 5971 Series mass selective detector (EI) 和 HP-INNOWax 层析柱 (长 30m, i. d. 0.25mm, 膜厚 0.25 微米) 上的 Hewlett Packard 5890 Series II 气相色谱。将产生的 1, 3-丙二醇的停留时间和质谱与真正的 1, 3-丙二醇的进行比较 (m/e: 57, 58)。

GC/MS 的一种替代方法需要对样品进行衍生化。在 1.0mL 的样品 (例如培养上清) 中加入 30μL 浓缩的高氯酸 (70% v/v)。混合后样品被冷冻并冰冻干燥。往冻干物质中加入 1: 1 混合的二 (三甲基硅氧烷基) 三氟乙酰胺: 吡啶, 剧烈混合并置于 65℃ 1h。通过离心沉淀不溶解物质使样品澄清。得到的液体分成两相, 上相用于分析。样品在 DB-5 层析柱 (48m, 0.25 I. D., 膜厚 0.25μM, 购自 J&W Scientific) 上进行层析, 将从培养物上清得到的 1, 3-丙二醇衍生物的停留时间和质谱与从真正的标准得到的衍生物的进行比较。TMS 衍生化的 1, 3-丙二醇含有的特征离子为 205, 177, 130 和 115AMU。

#### 维生素或辅酶 B<sub>12</sub> 的鉴定

将无细胞样品进行 HPLC 以定量辅酶 B<sub>12</sub> 和氰钴氨素。钴氨素的定量首先将 278nm 和 261nm 的峰面积比与标准的峰面积比进行比较, 然后将峰面积对应于钴氨素的标准曲线。

#### HPLC 方法

层析柱: Supelcosil LC-18-DB, 25cm×4.6mm (Supelco, Inc., Bellefonte, PA)  
Supelcosil LC-18-DB Precolumn kit

层析柱温度: 环境温度

样品室: 黑暗, 5℃

检测: 254nm, 和 360nm

注射体积: 25μL

流动相 A: 8.95g 乙酸钠. 3H<sub>2</sub>O  
5.88 mL 1.0M 四丁基铵盐氢氧化物 (TBAH)  
4L MQ H<sub>2</sub>O  
冰醋酸调 pH 至 4.6  
加入 210mL 流动相 B (见下文)

流动相 B: 4L MeOH  
5.88mL TBAH  
0.89ml 冰醋酸

梯度:

时间 (分钟)	流速 (mL/min)	A %	B %
0	1.0	100	0
3	1.0	75	25
9	1.0	60	40
11	1.0	0	100
13	1.0	0	100
15	1.0	100	0
15.5	0.1	100	0

### 编码甘油脱水酶基因 (dhaB) 以及编码 1, 3-丙二醇氧化还原酶 (dhaT) 基因的分离和克隆

dhaB 和 dhaT 的鉴定和分离方法基本上按照 US 5,686,276 进行, US 5,686,276 在此引入作为参考。在本发明中粘粒载体粘粒转化方法被用于从已知含有能将甘油加工成 1, 3-丙二醇的基因的细菌中克隆大片段染色体组 DNA。特别是, 用本领域中众所周知的方法分离了源

自肺炎克雷伯氏菌 ATCC 25955 的基因组 DNA 并用限制性内切酶 *Sau3A* 消化以插入粘粒载体 *Supercos1* 并用 *GigapackII* 包装抽提物包装。构建后用带有粘粒 DNA 的载体转化了数字大肠杆菌 XL1-Blue MR 细胞。通过使转化子在含甘油的情况下生长并且分析培养基中 1, 3-丙二醇的形成, 从而筛选转化子将甘油转化为 1, 3-丙二醇的能力。

分析了两个 1, 3-丙二醇阳性转化子, 并且它们的粘粒分别被命名为 pKP1 和 pKP2。DNA 测序表明和源自弗氏柠檬酸杆菌的甘油脱水酶基因 (*dhaB*) 具有很高的同源性, 表明这些转化子含有表明甘油脱水酶的基因。

源自 pKP1 的一个 12.1kb 的 *EcoRI-SalI* 片段被亚克隆到 pIBI31 (IBI Biosystem, New Havn, CN), 并测序和命名为 pHK28-26 (SEQ ID NO:10)。测序结果揭示了 *dha* 操纵子编码甘油脱水酶以及调节所必需的基因相关的开放阅读框的位点。参见 SEQ ID NO:10, *dhaK* (编码二羟基丙酮激酶) 开放阅读框一个片段被发现位于碱基 1-399; *dhaD* (编码甘油脱水酶) 的开放阅读框被发现位于碱基 983-2107; *dhaR* (编码抑制蛋白) 的开放阅读框被发现位于碱基 2209-4134; 开放阅读框 *dhaT* (编码 1, 3-丙二醇氧化还原酶) 被发现位于 5017-6180; 开放阅读框 *dhaB1* (编码甘油脱水酶 $\alpha$ 亚基) 被发现位于碱基 744-8711; 开放阅读框 *dhaB2* (编码甘油脱水酶 $\beta$ 亚基) 被发现位于 8724-9308; 开放阅读框 *dhaB3* (编码甘油脱水酶 $\gamma$ 亚基) 被发现位于 9311-9736; 开放阅读框 *dhaBX* (编码功能未知的蛋白) 被发现位于 9749-11572。此外, 开放阅读框 *orfY* (编码功能未知蛋白) 被发现位于 6202-6630; 开放阅读框 *orfX* (编码功能未知蛋白) 被发现位于 4643-4996; 开放阅读框 *orfW* (编码功能未知蛋白) 被发现位于 4112-4642。

#### 用于转化大肠杆菌的常用目的质粒的构建

##### 表达载体 pTacIQ 的构建

大肠杆菌表达载体 pTacIQ 是通过将 *lacIq* 基因 (Farabaugh, (1978), *Nature* 274(5673)765-769) 和 *tac* 启动子 (Amann et al., (1983), *Gene* 25, 167-178) 插入到 pBR322 (Sutcliffe, (1979), *Cold Spring Harb. Symp. Quant. Biol.* 43, 77-90) 的限制性内切酶位点 *EcoRI* 而构建的。一个多克隆位点和终止子序列取代了 pBR322 从 *EcoRI* 到 *SphI* 的序列。

### 甘油脱水酶基因 (dhaB1, 3, 3, X) 的亚克隆

利用在 5'端带有 EcoRI 位点和 3'端带有 XbaI 位点的引物 (SEQ ID NO: 12 和 SEQ ID NO: 13) 通过 PCR 从 pHK28-26 中克隆了 dhaB3 基因的开放阅读框。产物被亚克隆到 pLitmus29 (New England Biolab, Inc., Beverly, MA) 中以产生含有 dhaB3 基因的质粒 pDHAB3。

用限制性内切酶 KpnI 和 EcoRI 将含有 dhaB 操纵子的 dhaB1, dhaB2, dhaB3 和 dhaBX 全部编码区域的区域克隆到 pBluescriptIIKS<sup>+</sup> (Stratagene, La Jolla, CA) 中以建立质粒 pM7。

通过用 ApaI 和 XbaI 消化质粒 pM7 除去了 dhaBX 基因, 纯化 5.9kb 的片段并连接到源自质粒 pDHAB3 的 325-bp ApaI-XbaI 片段以建立含有 dhaB1, dhaB2 和 dhaB3 的 pM11。

利用在 5'端带有 HindIII 位点和共有核糖体结合位点 (consensus ribosome binding site), 3'端带有 XbaI 位点的引物 (SEQ ID NO: 14 和 SEQ ID NO: 15) 通过 PCR 从 pHK28-26 扩增了 dhaB1 基因的开放阅读框。扩增产物被亚克隆到 pLitmus28 (New England Biolab Inc.) 中以产生含 dhaB1 的质粒 pDT1。

含有一部分 dhaB1 基因的 NotI-XbaI 片段和 dhaB2 基因以及 dhaB3 基因被插入到 pDT1 以产生 dhaB 表达质粒, pDT2。pDT2 的含有 dhaB(1, 2, 3) 基因的 HindIII-XbaI 片段被插入到 pTacIQ 中以建立 pDT3。

### 1, 3-丙二醇脱水酶基因 (dhaT) 的亚克隆

pHK-28-26 的 KpnI-SacI 片段, 含有 1, 3-丙二醇脱水酶 (dhaT), 被亚克隆到 pBluescriptII KS<sup>+</sup> 中以建立质粒 pAH1。利用 pAH1 作为模板 DNA 和在 5'端带有 XbaI 位点, 3'端带有 BamHI 位点的合成引物 (SEQ ID NO: 16, SEQ ID NO: 17) 通过 PCR 扩增了 dhaT 基因。上述产物被亚克隆到 pCR-Script (Stratagene) 的 SrfI 位点以产生含有 dhaT 的质粒 pAH4 和 pAH5。质粒 pAH4 在正向含有 dhaT 基因以在 pCR-Script 中从 lac 启动子表达, pAH5 在反向含有 dhaT 基因。源自 pAH4 的含有 dhaT 基因的 XbaI-BamHI 片段被插入到 pTacIQ 中以产生质粒 pAH8。源自质粒 pAH8 的含有 RBS 和 dhaT 基因的 HindIII-BamHI 片段被插入到 pBluescriptIIKS<sup>+</sup> 以建立 pAH11。

### 构建 dhaT 和 dhaB(1, 2, 3) 表达盒

利用常规分子生物学方法从先前描述的单独的 dhaB(1, 2, 3) 和 dhaT 亚克隆装配了 dhaT 和 dhaB(1, 2, 3) 的表达盒。源自 pDT3 的含有 dhaB(1, 2, 3) 的 SpeI-SacI 片段被插入到 pAH11 的 SpeI-SacI 位点以建立 pAH24。SalI-XbaI 连接序列 (SEQ ID NO: 22 和 SEQ ID NO: 23) 被插入到用限制性内切酶 SalI-XbaI 消化了的 pAH5 中以建立 pDT16。该连接序列破坏了 XbaI 位点。然后源自 pDT16 的 1kb 的 SalI-MluI 片段被插入到 pAH24 中取代存在的 SalI-NotI 片段以建立 pDT18, 并且源自 pM7 的 NotI-XbaI 片段被插入到 pCL1920 (SEQ ID NO: 24)。源自 Streptomyces 的葡萄糖异构酶启动子序列 (SEQ ID NO: 25) 通过 PCR 被克隆并插入到 pLitmus28 的 EcoRI-HindIII 位点以构建 pDT5。通过将 pDT5 的 EcoRI-PvuII 片段插入到 pCL1920 的 EcoRI-PvuII 位点构建了 pCL1925 质粒。通过将 pDT21 的 HindIII-MluI 片段以及 pDT21 的 MluI-XbaI 片段克隆到 pCL1925 的 HindIII-XbaI 位点构建了 pDT24 质粒。

### 实施例 1

#### B<sub>12</sub> 转运基因表达盒的构建

构建了 B<sub>12</sub> 转运基因, btuB, btuC, btuD 和 btuE 的可作为复制元件存在的表达质粒。所有的表达质粒都用 trc 启动子进行转录。每个质粒还含有用于在含氨苄青霉素的培养基中筛选大肠杆菌的β-内酰胺酶的基因, 或用于在含氯霉素的培养基中进行筛选的氯霉素乙酰基转移酶基因。质粒的复制起始点为 ColE 或 p15A。

用在 5'端添加了 NcoI 位点和在 3'端添加了 BamHI 的引物 (SEQ ID NO: 18 及 SEQ ID NO: 19) 通过 PCR 从大肠杆菌染色体 DNA 扩增了 btuB 基因。反应混合物含有 10mM Tris, pH8.3, 50mM KCl, 1.5mM MgCl<sub>2</sub>, 0.0001%明胶, 200μM dATP, 200μM dCTP, 200μM dGTP, 200μM dTTP, 1μM 各种引物, 1-10ng 靶 DNA, 25 单位/mL Amplitaq™ DNA 聚合酶 (Perkin-Elmer Cetus, Norwalk CT)。PCR 参数是 94℃ 1min, 52℃ 1min, 72℃ 2min, 25 循环。在质粒 pTrc99A (Pharmacia, Piscataway, NJ) 的 NcoI 位点和 BamHI 位点之间克隆了 1905bp 的 PCR 产物以产生质粒 pBtuB1。质粒 pBtuB1 基因复制起始点 ColE1, 氨苄

青霉素抗性, lacIq 基因, btuB 从 Ptrc 开始表达。

为了构建质粒 pBtuB2, 从 pBtuB1 取出编码 lacIq, Ptrc 和 btuB 的 SphI/BamHI 片段并克隆到质粒 pACYC184 的 SphI/BamHI 位点。质粒 pBtuB2 具有复制起点 p15A, 氯霉素抗性和 lacIq 基因, 并且 btuB 从 PtrC 开始表达。

利用在 5'端添加了 BamHI 位点和在 3'端添加了 HindIII 的引物 (SEQID NO: 20 及 SEQ ID NO:219) 通过 PCR 从大肠杆菌染色体 DNA 扩增了 btuCED 基因。其 2557bp 的 PCR 产物被克隆到质粒 pACYC184 的 BamHI 和 HindIII 位点以产生质粒 pCED。质粒 pCED 基因复制起点 p15A, 和氯霉素抗性基因。

为了构建质粒 pBCED, 从 pBtuB1 取出编码 lacIq, Ptrc 和 btuB 的 SphI/BamHI 片段并克隆到质粒 pCED 的 SphI/BamHI 位点。质粒 pBtuB2 从 trc 启动子的下游开始依次具有复制起点 p15A, 氯霉素抗性, lacIq 基因, 和 btu 基因。

## 实施例 2

### 具有 B<sub>12</sub> 转运基因和 DhaB 活性的转化子

用 dha 质粒 pDT24 (specR), btuB 质粒 pBtuB1(ampR) 或 pBtuB2(ch1R), 或 btuBCED 质粒 pBCED(ch1R) 转化大肠杆菌菌株 FM5。筛选是在含 50mg/ml 壮观霉素, 50mg/ml 氨基青霉素或 50mg/ml 氯霉素的 LB 平板上进行的。抗适当抗生素的菌落被用于 1, 3-丙二醇生产和维生素 B<sub>12</sub> 摄取。

## 实施例 3

### 在转化了 pBCED 的 FM5 中辅酶 B<sub>12</sub> 的摄取提高

在含有 25mL 培养基 (培养基用 NH<sub>4</sub>OH 滴定至 pH6.8, 含有 0.2M KH<sub>2</sub>OH, 2.0g/L 柠檬酸, 2.0g/L mgSO<sub>4</sub>•7H<sub>2</sub>O, 1.2mL98% H<sub>2</sub>SO<sub>4</sub>, 0.3g/L 柠檬酸铵铁, 0.2g/L CaCl<sub>2</sub>•2H<sub>2</sub>O, 5mL 痕量金属混合物, 5g/L 酵母抽提物, 10g/LD-葡萄糖, 和适当的抗生素。痕量金属混合物含有 (g/L) : Na<sub>2</sub>SO<sub>4</sub>(4.0), MnSO<sub>4</sub>•H<sub>2</sub>O(0.80), ZnSO<sub>4</sub>•7H<sub>2</sub>O(1.6), CoSO<sub>4</sub>(0.52), CuSO<sub>4</sub>•7H<sub>2</sub>O(0.12) 和 FeSO<sub>4</sub>•7H<sub>2</sub>O(4.0) ) 的 250ml 三角烧瓶中 37℃ 过夜生长适当的菌株, 转速 250rpm。用 25mlM9 培养基稀释过夜培养物 (1/100) 并继续生长直至达到 OD<sub>660</sub> ~ 1.0。当加入 IPTG 时, 在此时间点将 IPTG 浓度添加至 0.2mM 并继续培养 1hr。

在该浓度往 M9 培养物中加入氰钴氨素（氰钴氨素，CNCbl）或辅酶 B<sub>12</sub>。所有涉及辅酶 B<sub>12</sub> 的操作都在暗处进行（红光）。加入钴氨素后立即取出 1ml 样品并沉淀细胞。然后继续在 250rpm 的转速下培养直至取出如下表 1 和表所列出的终点样品。

每 1mL 无细胞上清样品进行 HPLC 以定量钴氨素。钴氨素的定量首先将 278nm 和 261nm 的峰面积比与标准的峰面积比进行比较，然后将峰面积对应于钴氨素的标准曲线。

终点分析涉及从培养基中分离细胞，然后从细胞质中分离周质腔。方法基本上按照 Kaback (Methods of Enzymology, vol.22, pg. 99, 1971)。

回收的细胞沉淀称重，用 10mM Tris, pH8.0 洗两次。沉淀用 1g/80mL 30mM Tris, pH8.0/20%蔗糖从悬浮。在磁力搅拌板上搅拌时加入 EDTA 至 10mM，溶菌酶至 0.5mg/ml。这些悬浮液在室温搅拌 30min。溶菌酶/EDTA 温育后细胞结块并且如预期的那样沉淀。每份悬浮液在 15K rpm 4℃ 沉淀 20min。收集由稀释的质周腔组成的上清，记录体积并将样品进行 HPLC 分析。

回收的原生质体沉淀用 3ml 50mM 磷酸钾缓冲液 pH7.0 在组织匀浆器中匀浆。匀浆后加入 DNase 和 RNase 至 5mg/ml，并且悬浮液在 37℃ 水浴中温育。加入 EDTA 至 10mM，继续温育 15min。加入 MgSO<sub>4</sub> 至 15mM，继续温育 15min。

得到的悬浮液在 4℃ 39K 超速离心 1 小时。收集由稀释的细胞质组成的上清，记录体积，将样品继续 HPLC 分析。

根据假设 1μg 细胞（湿重）等于 1, 000, 000 细胞，细胞的体积是  $9 \times 10^{-13}$  mL，质周腔的体积为总细胞体积的 30% 计算质周腔和细胞质中的钴氨素浓度。

表 1  
pBtuB1A 对菌株 FM5 摄取 5μM 氰钴氨素的影响

菌株	时间 (hr)	质周腔	细胞质
FM5	16	6μM	6.5μM
FM5/pBtuB1	16	196μM	45.0μM

表 2  
pBCED 对菌株 FM5 摄取  $10\mu\text{M}$  辅酶  $\text{B}_{12}$  的影响

菌株	时间 (hr)	肉汤培养基	质周腔	细胞质
FM5/pBtuB2	0	$9.7\mu\text{M}$		
+IPTG	16	低于检测下限	$840\mu\text{M}$	$82\mu\text{M}$
FM5/pBCED	0	$10\mu\text{M}$		
+IPTG	16	低于检测下限	$280\mu\text{M}$	$170\mu\text{M}$

#### 实施例 4

转化了 pBCED 的 FM5/pDT24 中 1, 3-丙二醇的产生提高

在含有 25mL 培养基的 250ml 三角烧瓶中  $30^\circ\text{C}$  避光过夜培养大肠杆菌菌株 FM5/pDT24 和 FM5/pDT4/pBCED, 转速 250rpm. 培养基用  $\text{NH}_4\text{OH}$  滴定至 pH6.8, 含有 0.2M  $\text{KH}_2\text{OH}$ , 2.0g/L 柠檬酸, 2.0g/L  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 1.2mL98%  $\text{H}_2\text{SO}_4$ , 0.3g/L 柠檬酸铵铁, 0.2g/L  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 5mL 痕量金属混合物, 5g/L 酵母抽提物, 10g/LD-葡萄糖, 和适当的抗生素. 痕量金属混合物含有 (g/L):  $\text{Na}_2\text{SO}_4$ (4.0),  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ (0.80),  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ (1.6),  $\text{CoSO}_4$ (0.52),  $\text{CuSO}_4 \cdot 7\text{H}_2\text{O}$ (0.12) 和  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (4.0). 此外, pDT24 和 pBCED 分别需要  $50\mu\text{g/ml}$  壮观霉素和  $20\mu\text{g/ml}$  氯霉素.

FM5/pDT24 和 FM5/pDT4/pBCED 在如上文所描述的条件加上氰钴氨素, 羟基钴氨素 (羟基  $\text{B}_{12}$ ) 或辅酶  $\text{B}_{12}$  至终浓度  $0.40\mu\text{M}$  或  $4.0\mu\text{M}$ . 烧瓶中接种至起始 OD600 约 0.01AU, 通过加入 0.5N KOH 将 pH 保持在 6.2 以上, 通过加入 50% (W/W) 将葡萄糖浓度保持在 2g/L 以上. PH 用 ColorpHast strips(EM Science, Gibbstown, NJ) 检测. 葡萄糖浓度用 Trinder enzymatic assay(Sigma, St. Louis, MO) 检测. 在不同的时间取出小样以确定 3G 浓度(hplc 分析)和细胞密度(OD600). 结果在以下的表 3 和 4 出示.

表 3  
在 0.4 $\mu$ M 维生素，羟基，和辅酶 B<sub>12</sub> 存在的条件下  
pBCED 对 1, 3-丙二醇生产的影响

加入的 B <sub>12</sub> (0.4 $\mu$ M)	时间 (hr)	FM5/pDT24		FM5/pDT4/pBCED	
		1, 3-丙 二醇(g/L)	OD600 (AU)	1, 3-丙 二醇	OD600 (AU)
氰钴氨素	0	0.0	0.1	0.0	0.0
氰钴氨素	9	0.0	6.3	0.3	6.9
氰钴氨素	11	0.0	9.4	1.0	10.0
氰钴氨素	12	0.0	9.7	1.0	9.9
氰钴氨素	14	0.0	11.6	1.2	12.5
氰钴氨素	17	0.0	19.4	1.2	19.3
氰钴氨素	19	0.0	24.8	1.2	24.1
氰钴氨素	33	0.0	41.5	0.9	46.5
羟基 B <sub>12</sub>	0	0.0	0.1	0.0	0.0
羟基 B <sub>12</sub>	9	0.1	6.2	1.0	6.1
羟基 B <sub>12</sub>	11	0.3	8.8	2.0	8.3
羟基 B <sub>12</sub>	12	0.3	9.7	2.2	9.1
羟基 B <sub>12</sub>	14	0.3	10.4	2.3	10.5
羟基 B <sub>12</sub>	17	0.4	17.3	2.3	15.8
羟基 B <sub>12</sub>	19	0.4	22.0	2.2	18.2
羟基 B <sub>12</sub>	33	0.2	41.5	1.5	35.8
辅酶 B <sub>12</sub>	0	0.0	0.0	0.0	0.0
辅酶 B <sub>12</sub>	9	1.7	6.9	1.3	6.2
辅酶 B <sub>12</sub>	11	2.0	10.1	2.4	9.2
辅酶 B <sub>12</sub>	12	2.1	10.1	3.1	9.7
辅酶 B <sub>12</sub>	14	3.0	12.2	3.3	10.9
辅酶 B <sub>12</sub>	17	2.5	17.4	2.8	17.4
辅酶 B <sub>12</sub>	19	2.3	22.2	3.2	21.1
辅酶 B <sub>12</sub>	33	1.8	46.7	2.4	48.7

表 4  
在 4.0 $\mu$ M 维生素，羧基，和辅酶 B<sub>12</sub> 存在的条件下  
pBCED 对 1, 3-丙二醇生产的影响

加入的 B <sub>12</sub>	时间 (hr)	FM5/pDT24		FM5/pDT24/pBCED	
		1, 3-丙二 醇 (g/L)	OD600 (AU)	1, 3-丙二 醇 (g/L)	OD600 (AU)
氰钴氨素	0	0.0	0.2	0.0	0.1
氰钴氨素	8	0.1	8.9	0.4	9.7
氰钴氨素	10	0.1	11.9	0.8	11.6
氰钴氨素	12	0.3	13.7	1.5	15.9
氰钴氨素	14	0.8	17.8	3.3	23.0
氰钴氨素	16	1.3	24.7	6.1	29.6
氰钴氨素	33	1.6	36.6	10.2	40.7
羧基 B <sub>12</sub>	0	0.0	0.1	0.0	0.1
羧基 B <sub>12</sub>	8	0.4	9.0	1.7	9.5
羧基 B <sub>12</sub>	10	1.3	11.5	2.7	12.2
羧基 B <sub>12</sub>	12	2.8	12.9	3.8	14.6
羧基 B <sub>12</sub>	14	4.2	16.1	5.4	18.9
羧基 B <sub>12</sub>	16	5.5	19.9	7.2	25.1
羧基 B <sub>12</sub>	33	7.3	49.1	13.1	43.8
辅酶 B <sub>12</sub>	0	0.0	0.1	0.0	0.2
辅酶 B <sub>12</sub>	8	2.4	8.2	2.2	8.7
辅酶 B <sub>12</sub>	10	3.7	10.3	3.3	11.5
辅酶 B <sub>12</sub>	12	5.0	12.6	4.2	13.2
辅酶 B <sub>12</sub>	14	5.5	14.2	5.7	16.5
辅酶 B <sub>12</sub>	16	7.4	16.6	7.3	20.4
辅酶 B <sub>12</sub>	33	11.3	46.2	12.7	48.6

## 序列表

## (1) 一般信息

## (i) 申请人

- (A) 通信人: E. I. DUPONT DE NEMOURS AND COMPANY
- (B) 街道: 10007 MARKET STREET
- (C) 城市: WILMINGTON
- (D) 州: 特拉华
- (E) 国家: 美国
- (F) 邮编: 19898
- (G) 电话: 302 - 892 - 8112
- (H) 传真: 302 - 773 - 0164
- (I) 电传: 6717325

(A) 地址: GenenCOR. International , Inc.

(B) 街道: 925 Page Mill Road

(C) 城市: Palo Alto

(D) 州: 加利福尼亚

(E) 国家: 美国

(F) 邮编: 94304 - 1013

(ii) 发明标题: 用含有维生素 B12 转运基因的重组生物体生产  
1,3-丙二醇的方法

(iii) 序列数: 25

(iv) 计算机可读形式:

- (A) 介质类型: 软盘, 3.5 英寸
- (B) 计算机: IBM PC 计算机
- (C) 操作系统: 微软视窗 97
- (D) 软件: 微软 OFFICE97

(v) 目前的申请数据

- (A) 申请号:
- (B) 归档日期:
- (C) 分类:

(vi) 优先的申请数据:

- (A) 申请号: 60/085,190

- (B) 归档日期: 1998年6月30号
- (C) 分类:
- (vii) 律师/代理人信息
- (A) 名称: FLOYD, LINDA AXAMETHY
- (B) 注册号: 33, 692
- (C) 参考/文档号: CL-1245-A
- (2) SEQ ID NO: 1 的信息
- (i) 序列特征
- (A) 长度: 1845 碱基对
- (B) 类型: 核酸
- (C) 链: 双链
- (D) 拓扑性质: 线性
- (ii) 分子类型: DNA (基因组)
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 1

```

ATGATTA AAA AAGCTTCCCT GCTGACGGCG TGTTCCGTCA CGGCATTTTC CGCTTGGGCA 60
CAGGATACCA GCCCGGATAC TCTCGTCGTT ACTGCTAACC GTTTTGAACA GCCGCGCAGC 120
ACTGTGCTTG CACCAACCAC CGTTGTGACC CGTCAGGATA TCGACCGCTG GCAGTCGACC 180
TCGGTCAATG ATGTGCTGCG CCGTCTTCCG GCGGTCGATA TCACCCAAA CGGCGGTTCA 240
GGTCAGCTCT CATCTATTTT TATTCGCGGT ACAAATGCCA GTCATGTGTT GGTGTTAATT 300
GATGGCGTAC GCCTGAATCT GCGGGGGGTG AGTGGTTCTG CCGACCTTAG CCAGTTCCTT 360
ATTGCGCTTG TCCAGCGTGT TGAATATATC CGTGGGCCGC GCTCCGCTGT TTATGGTTCC 420
GATGCAATAG GCGGGGTGGT GAATATCATC ACGACGCGCG ATGAACCCGG AACGGAAATT 480
TCAGGAGGGT GGGGAAGCAA TAGTTATCAG AACTATGATG TCTCTACGCA GCAACAACCTG 540
GGGATAAGA CACGGGTAAC GCTGTTGGGC GATTATGCCC ATACTCATGG TTATGATGTT 600
GTTGCCTATG GTAATACCGG AACGCAAGCG CAGACAGATA ACGATGGTTT TTTAAGTAAA 660
ACGCTTTATG GCGCGCTGGA GCATAACTTT ACTGATGCCT GGAGCGGCTT TGTGCGCGGC 720

```

TATGGCTATG ATAACCGTAC CAATTATGAC GCGTATTATT CTCCCGGTTT ACCGTTGCTC 780  
 GATACCCGTA AACTCTATAG CCAAAGTTGG GACGCCGGGC TCGCTATAA CGGCGAACTG 840  
 ATTAAATCAC AACTCATTAC CAGCTATAGC CATAGCAAAG ATTACAATA CGATCCCCAT 900  
 TATGGTCGTT ATGATTTCGTC GCGCAGCTC GATGAGATGA AGCAATACAC CGTCCAGTGG 960  
 GCAAACAATG TCATCGTTGG TCACGGTAGT ATTGGTGCGG GTGTCGACTG GCAGAAACAG 1020  
 ACTACGACGC CGGGTACAGG TTATGTTGAG GATGGATATG ATCAACGTAA TACCGGCATC 1080  
 TATCTGACCG GGCTGCAACA AGTCGGCGAT TTTACCTTTG AAGGCGCCAG ACGCAGTGAC 1140  
 GATAACTCAC AGTTTGGTGC TCATGGAACC TGGCAAACCA GCGCCGGTTG GGAATTCATC 1200  
 GAAGTTATC GCTTCATTGC TTCCTACGGG ACATCTTATA AGGCACCAA TCTGGGGCAA 1260  
  
 CTGTATGGCT TCTACGGAAA TCCGAATCTG GACCCGGAGA AAAGCAAACA GTGGGAAGGC 1320  
 GCGTTTGAAG GCTTAACCGC TGGGGTGAAC TGGCGTATTT CCGGATATCG TAACGATGTC 1380  
 AGTGAATTGA TCGATTATGA TGATCACACC CTGAAATATT ACAACGAAGG GAAAGCGCGG 1440  
 ATTAAGGGCG TCGAGGCGAC CGCCAATTTT GATACCGGAC CACTGACGCA TACTGTGAGT 1500  
 TATGATTATG TCGATGCGCG CAATGCGATT ACCGACACGC CGTTGTTACG CCGTGCTAAA 1560  
 CAGCAGGTGA AATACCAGCT CGACTGGCAG TTGTATGACT TCGACTGGGG TATTACTTAT 1620  
 CAGTATTTAG GCACTCGCTA TGATAAGGAT TACTCATCTT ATCCTTATCA AACCGTTAAA 1680  
 ATGGGCGGTG TGAGCTTGTG GGATCTTGGC GTTGGGTATC CGGTCACCTC TCACCTGACA 1740  
 GTTCGTGGTA AAATAGCCAA CCTGTTTCGAC AAAGATTATG AGACAGTCTA TGGCTACCAA 1800  
 ACTGCAGGAC GGAATACAC CTTGTCTGGC AGCTACACCT TCTGA 1845

## (2) SEQ ID NO: 2 的信息

### (i) 序列特征

- (A) 长度: 1844 碱基对
- (B) 类型: 核酸
- (C) 链: 双链
- (D) 拓扑性质: 线性

### (ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

## (xi) 序列描述: SEQ ID NO: 2

```
ATGATTAAAA AAGCTACGCT GCTGACGGCG TTCTCCGTCA CGGCCTTTTC CGCTTGGGCG 60
CAGGACACTA GCCCGGATAC CCTGGTTGTC ACCGCCAACC GTTTTCAGCA GCCGCGCAGC 120
GCGGTTCTGG CGCCCATTAC CATCGTGACG CGTCAGGATA TTGAACGCTG GCAATCGACC 180
TCCGTAAATG ATGTTCTGCG CCGTTTGCCT GGCCTCGATA TTGCGCAGAG CGGCGGCGCG 240
CGACAAAACCT CCTCCATTTT CATTGCGGGC ACCAACTCCA GCCATGTACT GGTATTGATT 300
GACGGCGTGC GTCTGAATTT AGCAGGCGTG AGCGGGTCCG CCGATCTCAG CCAGTTCCCG 360
GTGTCGCTGG TACAGCGTAT TGAATATATA CGCGGTCCGC CCTCCGCTAT TTATGGTTCC 420
GATGCTATCG GCGGCGTAGT GAATATCATT ACGACGCGCG ATAACCCAGG CACAGAATTA 480
ACCGCTGGAT GGGGAAGCAA TAGCTACCAG AATTACGACA TCTCGACGCA ACAGCAACTT 540
GGCGAAATCA CGCGGGCGAC GTTGATCGGC GATTACGAAT ACACCAAAGG GTTTGACGTG 600
GTAGCGAAAG GCGGTACCGG GATGCAGGCG CAGCCTGACC GGGACGGCTT TTTGAGTAAA 660
```

```

ACGCTTTATG GCGCGTTAGA GCATACCTTT TCTGATCGCT GGAGCGGATT CGTGCGTGGT 720
TATGGCTACG ATAACCGTAC CGATTACGAC GCCTATTACT CGCCGGGCTC GCCGCTGATT 780
GATACACGCA AACTTTATAG CCAAAGCTGG GACGCCGGGC TGCACTTTAA TGGCGAAAGT 840
ATTCAGTCTC AGCTGGTTTC AAGCTATAGC CACAGTAAAG ATTACAAC TA TGATCCGCAC 900
TATGGCCGGT ATGATACCTC CGCCACGCTG GATGAGATGA AACAGTACAA TGTTC AATGG 960
ACCAACAGTG TGGTCGTGGG GACGGTAATG TTGGGGCGGG CGTAGACTGG CAGAAACAGA 1020
CTACCACGCC AGGTACCGGC TATGTGCCCC AGGGATATGA CCAGCGTAAT ACCGGGGTTT 1080
ACCTGACAGG ATTACAACAG TTGGGTGACT TCACTCTGGA AGCGGCGGCG CGCAGTGATG 1140
ACAAC TCCA GTTTGGTCGT CATGGTACAT GGCAAACCAG CGCGGGATGG GAGTTTATAG 1200
AAGGTTATCG CTTTATTGCC TCCTACGGAA CCTCCTACAA AGCGCCTAAT TTGGGCCAAC 1260
TGATGGTTA TTACGGTAAT CCGAACCTGA ATCCTGAAAA GAGTAAACAG TGGGAAGGCG 1320
CATTGAAGG GCTAACCGCT GGCCTCAGCT GGCCTATTTT AGGTTATCGT AACGATATTA 1380
ATGACATGAT CGATTATGAC GATCATCTGC AAAAATATTA CAACGAAGGT AAGGCGCGCA 1440
TTAAAGGTAT TGAGGCGACG GCGAATTTTC ATACCGGACC GTTAACGCAT ACGGTCAGTT 1500
ATGATTACGT TGATGCGCGT AATGCGATTA CCGATACGCC ATTACCCCGG CGTTCCAAAC 1560
AGATGGCAAA ATATCAACTT GACTGGGACG TTTACGATTT TGA CTGGGGG ATGACATATC 1620
AATACCTTGG TTCCCGCTAT GATTCGGATT ACTCCGCTTA CCCATACCGG ACAGTAAAAA 1680
TGGGCGGCGT CAGTTTATGG GATCTTACGG TTGCATATCC GGTCACCTCA CATCTGACAG 1740
TTCGTGGTAA AATAGCCAAC CTGTTTCGACA AAGATTACGA GACAGTTTAT GGCTACCAAA 1800
CTGCAGGACG AGAATACACC TTGTCTGGCA GCTACACCTT CTGA 1844

```

## (2) SEQ ID NO: 3 的信息

### (i) 序列特征

(A) 长度: 981 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 3

```

ATGCTGACAC TTGCCCGCCA ACAACAGCGA CAAAATATTC GCTGGTTATT ATGCCTGTCA    60

GTTTTGATGC TGCTGGCGCT TCTCTTAAGC CTTTGCGCCG GTGAACAATG GATCTCGCCA    120

GGTGA CTGGT TTACTCCTCG TGGCGAACTG TTCGTCTGGC AAATTCGCCT GCCACGTACG    180

CTGGCTGTAT TGCTGGTTGG TGC GGCGCTG GCTATATCCG GCGCTGTAAT GCAGGCGTTG    240

TTTGAAAATC CTCTGGCAGA ACCTGGACTA CTTGGCGTCT CTAACGGCGC AGGCGTGGGG    300

CTTATCGCCG CGGTATTGCT TGGGCAAGGG CTAACTCCCA ACTGGGCGCT AGGGCTGTGT    360

GCGATTCTGT GCGCGCTTAT CATCACTTTA ATACTCTTAC GTTTCGCCCG TCGTCATCTT    420

TCGACCAGTC GGTATTGCT GGCTGGCGTT GCATTAGGGA TTATCTGTAG CGCACTAATG    480

ACGTGGGCTA TCTACTTTTC CACCTCAGTT GATTTGCGTC AGCTGATGTA CTGGATGATG    540

GGCGGTTTTG GCGGCGTAGA CTGGCGGCAA AGCTGGCTGA TGCTGGCATT GATCCCCGTG    600

TTGTTGTGGA TCTGTTGTCA GTCCAGGCCG ATGAATATGT TAGCACTTGG CGAGATCTCG    660

GCGCGGCAAC TGGGTTTACC CCTGTGGTTC TGGCGCAATG TGCTGGTGGC AGCGACCGGC    720

TGGATGGTTG GCGTCAGTGT GCGCTGGCG GGTGCTATCG GCTTTATTGG TCTGGTGATC    780

CCCCATATTC TCCGGTTGTG TGTTTAAACC GATCATCGCG TATTACTTCC CGGCTGCGCG    840

CTGGCAGGGG CGAGCGCATT GCTGCTGGCC GATATTGTAG CGCGCCTGGC ATTAGCTGCC    900

GCAGAGCTGC CTATTGGCGT GGTCACCGCA ACGTTAGGTG CGCCGGTGTT TATCTGGTTA    960

TTGTTAAAAG CAGGACGTTA G                                                    981

```

## (2) SEQ ID NO: 4 的信息

### (i) 序列特征

(A) 长度: 750 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 4

```

ATGTCTATTG TGATGCAGTT ACAAGATGTT GCGGAATCTA CCCGCCTGGG GCCGCTTTCT 60
GGCGAGGTTT GGGCTGGGGA GATCCTGCAC CTGGTGGGGC CGAATGGCGC GGGTAAGAGT 120
ACCTTACTGG CGCGAATGGC CGGAATGACC AGCGGTAAGG GAAGCATTCA GTTCGCGGGG 180
CAACCACTGG AAGCATGGTC CGCAACAAAA CTCGCGCTGC ATCGCGCCTA TCTTTCACAA 240
CAGCAGACGC CGCCGTTTGC AACGCCGGTC TGGCACTACC TGACACTGCA TCAGCACGAT 300

AAAACGCGTA CCGAACTACT GAATGATGTC GCAGGGGCGC TGGCTCTGA TGACAAACTC 360
GGACGTAGCA CCAATCAACT TTCCGGCGGT GAATGGCAAC GCGTACGTCT TGCTGCGGTG 420
GTGTTGCAAA TCACACCACA AGCCAATCCC GCAGGCCAAT TGCTGCTTCT TGATGAGCCG 480
ATGAACAGTC TTGATGTTGC GCAACAAAGT GCGTTAGACA AAATTCTGAG CGCGCTGTGT 540
CAGCAAGGAC TGGCGATTGT GATGAGCAGT CACGATCTCA ACCACACATT GCGTCATGCC 600
CATCGGGCGT GGTGCTAAA AGGTGGAAAA ATGCTGGCCA GTGGACGCAG GGAAGAGGTC 660
CTCACGCCGC CAAATCTGGC GCAGGCCTAT GGGATGAATT TTCGCCGTCT GGATATCGAA 720
GGTCACAGAA TGCTGATTTC GACCATCTGA 750

```

## (2) SEQ ID NO: 5 的信息

### (i) 序列特征

(A) 长度: 552 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 5

```

ATGCAAGATT CCATTCTGAC GACCGTAGTG AAAGATATCG ACGGTGAAGT GACCACGCTG   60
GAGAAGTTTCG CCGGTAATGT GCTGTTGATT GTCAATGTTCG CCTCAAAGTG TGGCTTAACG  120
CCGCAATATG AGCAGTTGGA GAATATTCAG AAAGCCTGGG TCGATCGAGG TTTTATGGTG  180
CTGGGATTCC CGTGCAACCA GTTTCTGGAA CAAGAACCGG GCAGCGATGA AGAGATTAAA  240
ACTTACTGTA CCACCACATG GGGGGTGACG TTCCCGATGT TCAGTAAGAT TGAAGTTAAT  300
GGCGAAGGAC GCCATCCGCT GTATCAAAAA TTGATTGCCG CAGCGCCGAC CGCAGTCGCG  360
CCGGAAGAGA GCGGATTCTA TGCCCGTATG GTCAGCAAAG GCCGTGCACC GCTGTACCCG  420
GATGATATTT TATGGAATTT TGAAAAATTC CTGGTTGGCA GGGACGGAAA AGTCATCCAG  480
CGTTTTTCCC CGGATATGAC GCCGGAAGAT CCCATTGTGA TGGAAAGCAT TAAACTGGCG  540
TTGGCAAAAT AA                                                                552

```

## (2) SEQ ID NO: 6 的信息

### (i) 序列特征

(A) 长度: 1668 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 6

ATGAAAAGAT CAAAACGATT TGCAGTACTG GCCCAGCGCC CCGTCAATCA GGACGGGCTG 60  
 ATGGCGAGT GGCCTGAAGA GGGGCTGATC GCCATGGACA GCCCCTTTGA CCCGGTCTCT 120  
 TCAGTAAAAG TGGACAACGG TCTGATCGTC GAACTGGACG GCAAACGCCG GGACCAGTTT 180  
 GACATGATCG ACCGATTTAT CGCCGATTAC GCGATCAACG TTGAGCGCAC AGAGCAGGCA 240  
 ATGCGCCTGG AGGCGGTGGA AATAGCCCGT ATGCTGGTGG ATATTCACGT CAGCCGGGAG 300  
 GAGATCATTG CCATCACTAC CGCCATCACG CCGGCCAAAG CGGTCGAGGT GATGGCGCAG 360  
 ATGAACGTGG TGGAGATGAT GATGGCGCTG CAGAAGATGC GTGCCCGCCG GACCCCTCC 420  
 AACCAGTGCC ACGTCACCAA TCTCAAAGAT AATCCGGTGC AGATTGCCGC TGACGCCGCC 480  
 GAGGCCGGGA TCCGCGGCTT CTCAGAACAG GAGACCACGG TCGGTATCGC GCGCTACCGC 540  
 CCGTTTAACG CCCTGGCGCT GTTGGTCGGT TCGCAGTGC GCGCCCCGG CGTGTGACG 600  
 CAGTGCTCGG TGGAAAGAGC CACCGAGCTG GAGCTGGGCA TCGGTGGCTT AACCAGCTAC 660  
 GCCGAGACGG TGTCGGTCTA CGGCACCGAA GCGGTATTTA CCGACGGCGA TGATACGCCG 720  
 TGGTCAAAGG CGTTCCTCGC CTCGGCCTAC GCCTCCCGCG GGTGAAAAT GCGCTACACC 780  
 TCCGGCACCG GATCCGAAGC GCTGATGGGC TATTCGGAGA GCAAGTCGAT GCTCTACCTC 840  
 GAATCGCGCT GCATCTTCAT TACTAAAGGC GCCGGGTTC AGGGACTGCA AAACGGCGCG 900  
 GTGAGCTGTA TCGGCATGAC CGGCCTGTG CCGTCGGGCA TTCGGGCGGT GCTGGCGGAA 960  
 AACCTGATCG CCTCTATGCT CGACCTCGAA GTGGCGTCCG CCAACGACCA GACTTTCTCC 1020  
 CACTCGGATA TTCGCCGCAC CGCGCGCACC CTGATGCAGA TGCTGCCGGG CACCGACTTT 1080  
 ATTTTCTCCG GCTACAGCGC GGTGCCGAAC TACGACAACA TGTTCCGGCG CTCGAACTTC 1140  
 GATGCGGAAG ATTTTGATGA TTACAACATC CTGCAGCGTG ACCTGATGGT TGACGGCGGC 1200  
 CTGCGTCCGG TGACCGAGGC GGAAACCATT GCCATTCGCC AGAAAGCGGC GCGGGCGATC 1260  
 CAGGCGGTTT TCCGCGAGCT GGGGCTGCCG CCAATCGCCG ACGAGGAGGT GGAGGCCGCC 1320  
 ACCTACGCGC ACGGCAGCAA CGAGATGCCG CCGCGTAACG TGGTGGAGGA TCTGAGTGCG 1380  
 GTGGAAGAGA TGATGAAGCG CAACATCACC GGCCTCGATA TTGTCGGCGC GCTGAGCCGC 1440  
  
 AGCGGCTTTG AGGATATCGC CAGCAATATT CTCAATATGC TGCACCAGCG GGTACCAGGC 1500  
 GATTACCTGC AGACCTCGGC CATTCTCGAT CGGCAGTTCG AGGTGGTGAG TCGGGTCAAC 1560  
 GACATCAATG ACTATCAGGG GCCGGGCACC GGCTATCGCA TCTCTGCCGA ACGCTGGGCG 1620  
 GAGATCAAAA ATATTCGGG CGTGGTTCAG CCCGACACCA TTGAATAA 1668

## (2) SEQ ID NO: 7 的信息

## (i) 序列特征

(A) 长度: 585 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 7

```

GTGCAACAGA CAACCCAAAT TCAGCCCTCT TTTACCCTGA AAACCCGCGA GGGCGGGGTA 60
GCTTCTGCCG ATGAACGCGC CGATGAAGTG GTGATCGGCG TCGGCCCTGC CTTCGATAAA 120
CACCAGCATC ACACTCTGAT CGATATGCCC CATGGCGCGA TCCTCAAAGA GCTGATTGCC 180
GGGGTGGAA GAGAGGGGCT TCACGCCCCG GTGGTGCGCA TTCTGCGCAC GTCCGACGTC 240
TCCTTTATGG CCTGGGATGC GGCCAACCTG AGCGGCTCGG GGATCGGCAT CGGTATCCAG 300
TCGAAGGGGA CCACGGTCAT CCATCAGCGC GATCTGCTGC CGCTCAGCAA CCTGGAGCTG 360
TTCTCCCAGG CGCCGCTGCT GACGCTGGAG ACCTACCGGC AGATTGGCAA AAACGCTGCG 420
CGCTATGCGC GCAAAGAGTC ACCTTCGCCG GTGCCGGTGG TGAACGATCA GATGGTGC GG 480
CCGAAATTTA TGGCCAAAGC CGCGCTATTT CATATCAAAG AGACCAAACA TGTGGTGCAG 540
GACGCCGAGC CCGTCACCCT GCACATCGAC TTAGTAAGGG AGTGA 585

```

## (2) SEQ ID NO: 8 的信息

## (i) 序列特征

(A) 长度: 426 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

## (xi) 序列描述: SEQ ID NO: 8

```

ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG CCCGGAGCAT    60
ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT GCTCTCTGGC    120
GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA GCGCAGATT    180
GCCGAGCAGA TGCAGCGCCA TGCGGTGGCG CGCAATTTCC GCCGCGCGGC GGAGCTTATC    240
GCCATTCCCTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TCGGCCCGTT CCGCTCCTCG    300
CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC GACAGTGAAT    360
GCCGCCTTTG TCCGGGAGTC GCGGGAAGTG TATCAGCAGC GGCATAAGCT GCGTAAAGGA    420
AGCTAA                                           426

```

## (2) SEQ ID NO: 9 的信息

## (i) 序列特征

(A) 长度: 1164 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 9

```

ATGAGCTATC GTATGTTTGA TTATCTGGTG CCAAACGTTA ACTTTTTTGG CCCCAACGCC   60
ATTTCCGTAG TCGGCGAACG CTGCCAGCTG CTGGGGGGGA AAAAAGCCCT GCTGGTCACC  120
GACAAAGGCC TCGGGGCAAT TAAAGATGGC GCGGTGGACA AAACCCTGCA TTATCTGCGG  180
GAGGCCGGGA TCGAGGTGGC GATCTTTGAC GGCCTCGAGC CGAACCCGAA AGACACCAAC  240
GTGCGCGACG GCCTCGCCGT GTTTCGCCGC GAACAGTGCG ACATCATCGT CACCGTGGGC  300
GGCGGCAGCC CGCACGATTG CGGCAAAGGC ATCGGCATCG CCGCCACCCA TGAGGGCGAT  360
CTGTACCAGT ATGCCGGAAT CGAGACCCTG ACCAACCCGC TGCCGCCTAT CGTCGCGGTC  420
AATACCACCG CCGGCACCGC CAGCGAGGTC ACCCGCCACT GCGTCCTGAC CAACACCGAA  480
ACCAAAGTGA AGTTTGTGAT CGTCAGCTGG CGCAAAGTGC CGTCGGTCTC TATCAACGAT  540
CCTACTGCTG TGATCGGTAA ACCGGCCGCC CTGACCGCGG CGACCGGGAT GGATGCCCTG  600
ACCCACGCCG TAGAGGCCTA TATCTCCAAA GACGCTAACC CGGTGACGGA CGCCGCCGCC  660
ATGCAGGCGA TCCGCCTCAT CGCCCCCAAC CTGCGCCAGG CCGTGGCCCT CGGCAGCAAT  720

CTGCAGGCGC GGGAAAACAT GGCCTATGCT TCTCTGCTGG CCGGGATGGC TTTCAATAAC  780
GCCAACCTCG GCTACGTGCA CGCCATGGCG CACCAGCTGG GCGGCCTGTA CGACATGCCG  840
CACGGCGTGG CCAACGCTGT CCTGCTGCCG CATGTGGCGC GCTACAACCT GATCGCCAAC  900
CCGGAGAAAT TCGCCGATAT CGCTGAACTG ATGGGCGAAA ATATCACCGG ACTGTCCACT  960
CTCGACGCGG CGGAAAAGC CATCGCCGCT ATCACGCGTC TGTCGATGGA TATCGGTATT 1020
CGCAGCATC  TCGCGATCT  GGGGGTAAA  GAGGCCGACT  TCCCCTACAT  GCGGAGATG 1080
GCTCTAAAAG ACGGCAATGC GTTCTCGAAC CCGCGTAAAG GCAACGAGCA GGAGATTGCC 1140
GCGATTTTCC GCCAGGCATT CTGA                                     1164

```

## (2) SEQ ID NO: 10 的信息

### (i) 序列特征

- (A) 长度: 12145 碱基对
- (B) 类型: 核酸
- (C) 链: 双链
- (D) 拓扑性质: 线性

### (ii) 分子类型: DNA (基因组)

### (iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 10

GTCGACCACC	ACGGTGGTGA	CTTTAATGCC	GCTCTCATGC	AGCAGCTCGG	TGGCGGTCTC	60
AAAATTCAGG	ATGTCGCCGG	TATAGTTTTT	GATAATCAGC	AAGACGCCTT	CGCCGCCGTC	120
AATTTCATC	GCGCATTCAA	ACATTTTGTG	CGGCGTCGGC	GAGGTGAATA	TTTCCCCCGG	180
ACAGGCGCCG	GAGAGCATGC	CCTGGCCGAT	ATAGCCGCAG	TGCATCGGTT	CATGTCCGCT	240
GCCGCCGCCG	GAGAGCAGGG	CCACCTTGCC	AGCCACCGGC	GCGTCGGTGC	GGGTCACATA	300
CAGCGGGTCC	TGATGCAGGG	TCAGCTGCGG	ATGGGCTTTA	GCCAGCCCCT	GTAATTGTTC	360
ATTCAGTACA	TCTTCAACAC	GGTTAATCAG	CTTTTTTATT	ATTCAGTGCT	CCGTTGGAGA	420
AGGTTTCGATG	CCGCCTCTCT	GCTGGCGGAG	GCGGTCATCG	CGTAGGGGTA	TCGTCTGACG	480
GTGGAGCGTG	CCTGGCGATA	TGATGATTCT	GGCTGAGCGG	ACGAAAAAAAA	GAATGCCCCG	540
ACGATCGGGT	TTCATTACGA	AACATTGCTT	CCTGATTTTG	TTTCTTTATG	GAACGTTTTT	600
GCTGAGGATA	TGGTGAAAAT	GCGAGCTGGC	GCGCTTTTTT	TCTTCTGCCA	TAAGCGGCGG	660
TCAGGATAGC	CGGCGAAGCG	GGTGGGAAAA	AATTTTTTGC	TGATTTTCTG	CCGACTGCGG	720
GAGAAAAGGC	GGTCAAACAC	GGAGGATTGT	AAGGGCATTG	TGCGGCAAAG	GAGCGGATCG	780

```

GGATCGCAAT CCTGACAGAG ACTAGGGTTT TTTGTTCCAA TATGGAACGT AAAAAATTAA      840
CCTGTGTTTC ATATCAGAAC AAAAAGGCGA AAGATTTTTT TGTTCCCTGC CGGCCCTACA      900
GTGATCGCAC TGCTCCGGTA CGCTCCGTTT AGGCCGCGCT TCACTGGCCG GCGCGGATAA      960
CGCCAGGGCT CATCATGTCT ACATGCGCAC TTATTTGAGG GTGAAAGGAA TGCTAAAAGT     1020
TATTCAATCT CCAGCCAAAT ATCTTCAGGG TCCTGATGCT GCTGTTCTGT TCGGTCAATA     1080
TGCCAAAAAC CTGGCGGAGA GCTTCTTCGT CATCGCTGAC GATTTCGTAA TGAAGCTGGC     1140
GGGAGAGAAA GTGGTGAATG GCCTGCAGAG CCACGATATT CGCTGCCATG CGGAACGGTT     1200
TAACGGCGAA TGCAGCCATG CGGAAATCAA CCGTCTGATG GCGATTTTGC AAAAAACAGG     1260
CTGCCCGGCC GTGGTCGGGA TCGGCGGTGG TAAAACCCTC GATACCGCGA AGGCGATCGG     1320
TTACTACCAG AAGTGCCCGG TGGTGGTGAT CCCGACCATC GCCTCGACCG ATGCGCCAAC     1380
CAGCGCGCTG TCGGTGATCT ACACCGAAGC GGGCGAGTTT GAAGAGTATC TGATCTATCC     1440
GAAAAACCCG GATATGGTGG TGATGGACAC GGCATTATC GCCAAAGCGC CGGTACGCCT     1500
GCTGGTCTCC GGCATGGGCG ATGCGCTCTC CACCTGGTTC GAGGCCAAAG CTTGCTACGA     1560
TGCGCGCGCC ACCAGCATGG CCGGAGGACA GTCCACCGAG GCGGCGCTGA GCCTCGCCCG     1620
CCTGTGCTAT GATACGCTGC TGGCGGAGGG CGAAAAGGCC CGTCTGGCGG CGCAGGCCGG     1680
GGTAGTGACC GAAGCGCTGG AGCGCATCAT CGAGGCGAAC ACTTACCTCA GCGGCATTGG     1740
CTTTGAAAGC AGTGGCCTGG CCGCTGCCCA TGCAATCCAC AACGGTTTCA CCATTCTTGA     1800
AGAGTGCCAT CACCTGTATC ACGGTGAGAA AGTGGCCTTC GGTACCCTGG CGCAGCTGGT     1860
GCTGCAGAAC AGCCCGATGG ACGAGATTGA AACGGTGCA GGGTTCTGCC AGCGCGTCGG     1920
CCTGCCGGTG ACGCTCGGCG AGATGGGCGT CAAAGAGGGG ATCGACGAGA AAATCGCCGC     1980
GGTGGCGAAA GCTACCTGCG CGGAAGGGGA AACCATCCAT AATATGCCGT TTGCGGTGAC     2040
CCCGGAGAGC GTCCATGCCG CTATCCTCAC CGCCGATCTG TTAGGCCAGC AGTGGCTGGC     2100
GCGTTAATTC GCGGTGGCTA AACCGCTGGC CCAGGTCAGC GGTTTTTTCTT TCTCCCTCC     2160
GGCAGTCGCT GCCGGAGGGG TTCTCTATGG TACAACGCGG AAAAGGATAT GACTGTTTCAG     2220
ACTCAGGATA CCGGGAAGGC GGTCTCTTCC GTCATTGCC AGTCATGGCA CCGCTGCAGC     2280
AAGTTTATGC AGCGCGAAAC CTGGCAAACG CCGCACCAGG CCCAGGGCCT GACCTTCGAC     2340
TCCATCTGTC GGCCTAAAAC CGCGCTGCTC ACCATCGGCC AGGCGGCGCT GGAAGACGCC     2400
TGGGAGTTTA TGGACGGCCG CCCCTGCGCG CTGTTTATTC TTGATGAGTC CGCCTGCATC     2460
CTGAGCCGTT GCGGCGAGCC GCAAACCCTG GCCCAGCTGG CTGCCCTGGG ATTTGCGGAC     2520

```

GGCAGCTATT GTGCGGAGAG CATTATCGGC ACCTGCGCGC TGTCGCTGGC CGCGATGCAG 2580  
GGCCAGCCGA TCAACACCGC CGGCGATCGG CATTTTAAGC AGGCGCTACA GCCATGGAGT 2640  
TTTTGCTCGA CGCCGGTGTG TGATAACCAC GGGCGGCTGT TCGGCTCTAT CTCGCTTTGC 2700  
TGCTGTGTCG AGCACCAGTC CAGCGCCGAC CTCTCCCTGA CGCTGGCCAT CGCCCGCGAG 2760  
GTGGGTAACT CCCTGCTTAC CGACAGCCTG CTGGCGGAAT CCAACCGTCA CCTCAATCAG 2820  
ATGTACGGCC TGCTGGAGAG CATGGACGAT GGGGTGATGG CGTGGAACGA ACAGGGCGTG 2880  
CTGCAGTTTC TCAATGTTCA GCGGCGGAGA CTGCTGCATC TTGATGCTCA GGCCAGCCAG 2940  
GGGAAAATA TCGCCGATCT GGTGACCCTC CCGGCGCTGC TCGCGCGCGC CATCAAACAC 3000  
GCCCCGCGCC TGAATCACGT CGAAGTCACC TTTGAAAGTC AGCATCAGTT TGTCCGATGCG 3060  
GTGATCACCT TAAAACCGAT TGTCGAGGCG CAAGGCAACA GTTTTATTCT GCTGCTGCAT 3120  
CCGGTGGAGC AGATGCGGCA GCTGATGACC AGCCAGCTCG GTAAAGTCAG CCACACCTTT 3180  
GAGCAGATGT CTGCCGACGA TCCGGAACC CGACGCCTGA TCCACTTTGG CCGCCAGGCG 3240  
GCGCGCGGCG GCTTCCCGGT GCTACTGTGC GCGAAGAGG GGGTCGGGAA AGAGCTGCTG 3300  
AGCCAGGCTA TTCACAATGA AAGCGAACGG GCGGGCGGCC CCTACATCTC CGTCAACTGC 3360  
CAGCTATATG CCGACAGCGT GCTGGGCCAG GACTTTATGG GCAGCGCCCC TACCGACGAT 3420  
GAAAATGGTC GCCTGAGCCG CCTTGAGCTG GCCAACGGCG GCACCCTGTT TCTGGAAAAG 3480  
ATCGAGTATC TGGCCCGGA GCTGCAGTCG GCTCTGCTGC AGGTGATTAA GCAGGGCGTG 3540  
CTCACCCGCC TCGACGCCCC GCGCCTGATC CCGGTGGATG TGAAGGTGAT TGCCACCACC 3600  
ACCGTCGATC TGCCCAATCT GGTGGAACAG AACCGCTTTA GCCGCCAGCT GTACTATGCG 3660  
CTGCACTCCT TTGAGATCGT CATCCCGCCG CTGCGCGCCC GACGCAACAG TATTCCGTCG 3720  
CTGGTGCATA ACCGGTTGAA GAGCCTGGAG AAGCGTTTCT CTTCGCGACT GAAAGTGGAC 3780  
GATGACGCGC TGGCACAGCT GGTGGCCTAC TCGTGGCCGG GGAATGATTT TGAGCTCAAC 3840  
AGCGTCATTG AGAATATCGC CATCAGCAGC GACAACGGCC ACATTCCGCT GAGTAATCTG 3900  
CCGGAATATC TCTTTTCCGA GCGGCCGGG GGGGATAGCG CGTCATCGCT GCTGCCGGCC 3960  
AGCCTGACTT TTAGCGCCAT CGAAAAGGAA GCTATTATTC ACGCCGCCCG GGTGACCAGC 4020  
GGGCGGGTGC AGGAGATGTC GCAGCTGCTC AATATCGGCC GCACCACCCT GTGGCGCAA 4080  
ATGAAGCAGT ACGATATTGA CGCCAGCCAG TTCAAGCGCA AGCATCAGGC CTAGTCTCTT 4140  
CGATTGCGCG CATGGAGAAC AGGGCATCCG ACAGGCGATT GCTGTAGCGT TTGAGCGCGT 4200  
CGCGCAGCGG ATGCGCGCGG TCCATGGCCG TCAGCAGGCG TTCGAGCCGA CGGGACTGGG 4260

TGC GCGCCAC GTGCAGCTGG GCAGAGGCGA GATTCCTCCC CGGGATCAGG AACTGTTTTA 4320  
 ACGGGCCGCT CTCGGCCATA TTGCGGTGCA TAAGCCGCTC CAGGGCGGTG ATCTCCTCTT 4380  
 CGCCGATCGT CTGGCTCAGG CGGGTCAGGC CCCGCGCATC GCTGGCCAGT TCAGCCCCCA 4440  
 GCACGAACAG CGTCTGCTGA ATATGGTGCA GGCTTTCCCG CAGCCCGGCG TCGCGGGTCG 4500  
 TGGCGTAGCA GACGCCCAGC TGGGATATCA GTTCATCGAC GGTGCCGTAG GCCTCGACGC 4560  
 GAATATGGTC TTTCTCGATG CGGCTGCCGC CGTACAGGGC GGTGGTGCCCT TTATCCCCGG 4620  
 TCGGGTATA GATACGATAC ATTCACTTTC TCTCACTTAA CGGCAGGACT TTAACCAGCT 4680  
 GCCCCGCGTT GCGCGCCAGC GTACGCAGTT GATCGTCGCT ATCGGTGACG TGTCCGGTAG 4740  
 CCAGCGGCCG GTCCCGCCGC AGCTGGGCAT GAGTGAGGGC TATCTCGCCG GACGCGCTGA 4800  
 GCCCATAACC CACCCGAGG GGCAGCTTC TGGCCGCCAG GCGCCAGC GCAGCGGCGT 4860  
 CACCGCCTCC GTCATAGGTT ATGGTCTGGC AGGGGACCCC CTGCTCCTCC AGCCCCAGC 4920  
 ACAGCTCATT GATGGCGCCG GCATGGTGCC CGCGCGGATC GTAAAACAGG CGTACGCCTG 4980  
 GCGGTGAAAAG CGACATGACG GTCCCCTCGT TAACACTCAG AATGCCTGGC GGAAAATCGC 5040  
 GGCAATCTCC TGCTCGTTGC CTTTACGCGG GTTCGAGAAC GCATTGCCGT CTTTTAGAGC 5100  
 CATCTCCGCC ATGTAGGGGA AGTCGGCCTC TTTTACCCCC AGATCGCGCA GATGCTGCGG 5160  
 AATACCGATA TCCATCGACA GACGCGTGAT AGCGGCGATG GCTTTTTCCG CCGCGTCGAG 5220  
 AGTGGACAGT CCGGTGATAT TTTGCCCCAT CAGTTCAGCG ATATCGGCGA ATTTCTCCGG 5280  
 GTTGGCGATC AGGTTGTAGC GCGCCACATG CGGCAGCAGG ACAGCGTTGG CCACGCCGTG 5340  
 CGGCATGTCT TACAGGCCGC CCAGCTGGTG CGCCATGGCG TGCACGTAGC CGAGGTTGGC 5400  
 GTTATTGAAA GCCATCCCGG CCAGCAGAGA AGCATAGGCC ATGTTTTCCC GCGCCTGCAG 5460  
 ATTGCTGCCG AGGGCCACGG CCTGGCGCAG GTTGCGGGCG ATGAGGCGGA TCGCCTGCAT 5520  
 GCGGCGGGCG TCCGTCACCG GGTTAGCGTC TTTGGAGATA TAGGCCTCTA CGGCGTGGGT 5580  
 CAGGGCATCC ATCCCGGTCTG CCGCGGTCAG GCGGCGGGT TTACCGATCA TCAGCAGTGG 5640  
 ATCGTTGATA GAGACCGACG GCAGTTTGCG CCAGCTGACG ATCACAACT TCACTTTGGT 5700  
 TTCGGTGTG GTCAGGACGC AGTGGCGGGT GACCTCGCTG GCGGTGCCGG CGGTGGTATT 5760  
 GACCGCGACG ATAGGCGGCA GCGGTTGGT CAGGGTCTCG ATTCCGGCAT ACTGGTACAG 5820  
 ATCGCCCTCA TGGGTGGCGG CGATGCCGAT GCCTTTGCCG CAATCGTGCG GGCTGCCGCC 5880  
 GCCACGGTG ACGATGATGT CGCACTGTTT GCGGCGAAAC ACGGCGAGGC CGTCGCGCAC 5940  
 GTTGGTGTCT TTCGGGTTCT GCTCGACGCC GTCAAAGATC GCCACCTCGA TCCCGGCCTC 6000

CCGCAGATAA TGCAGGGTTT TGTCCACCGC GCCATCTTTA ATTGCCCGCA GGCCTTTGTC 6060  
 GGTGACCAGC AGGGCTTTTT TCCCCCCAG CAGCTGGCAG CGTTCGCCGA CTACGGAAAT 6120  
 GGC GTTGGGG CCAAAAAAGT TAACGTTTGG CACCAGATAA TCAAACATAC GATAGCTCAT 6180  
 AATATACCTT CTCGCTTCAG GTTATAATGC GGAAAAACAA TCCAGGGCGC ACTGGGCTAA 6240  
 TAATTGATCC TGCTCGACCG TACCGCCGCT AACGCCGACG GCGCCAATTA CCTGCTCATT 6300  
 AAAAATAACT GGCAGGCCGC CGCCAAAAAT AATAATTCCG TGTGGTTGG TTAGCTGCAG 6360  
 ACCGTACAGA GATTGTCTCG GCTGGACCGC TGACGTAATT TCATGGGTAC CTTGCTTCAG 6420  
 GCTGCAGGCG CTCCAGGCTT TATTCAGGGA AATATCGCAG CTGGAGACGA AGGCCTCGTC 6480  
 CATCCGCTGG ATAAGCAGCG TGTTCCTCC GCGGTCAACT ACGGAAAACA CCACCGCCAC 6540  
 GTTGATCTCA GTGGCTTTTT TTTCCACCGC CGCCGCCATT TGCTGGGCGG CGGCCAGGGT 6600  
 GATTGTCTGA ACTTGTGGC TCTTGTTCAT CATTCTCTCC CGCACCAGGA TAACGCTGGC 6660  
 GCGAATAGTC AGTAGGGGGC GATAGTAAAA AACTATTACC ATTCCGTTGG CTTGCTTTAT 6720  
 TTTTGTACGC GTTATTTTGT CGCCCGCCAT GATTTAGTCA ATAGGGTTAA AATAGCGTCG 6780  
 AAAAACGTA ATTAAGGGCG TTTTTTATTA ATTGATTAT ATCATTGCGG GCGATCACAT 6840  
 TTTTTATTTT TGCCGCCGGA GTAAAGTTTC ATAGTGAAAC TGTCGGTAGA TTTCGTGTGC 6900  
 CAAATGAAA CGAAATTAAA TTTATTTTTT TCACCACTGG CTCATTTAAA GTTCCGCTAT 6960  
 TGCCGGTAAT GGCCGGGCGG CAACGACGCT GGCCGGCGT ATTCGCTACC GTCTGCGGAT 7020  
 TTCACCTTTT GAGCCGATGA ACAATGAAAA GATCAAAACG ATTTGCAGTA CTGGCCAGC 7080  
 GCCCCGTCAA TCAGGACGGG CTGATTGGCG AGTGGCCTGA AGAGGGGCTG ATCGCCATGG 7140  
 ACAGCCCCTT TGACCCGGTC TCTTCAGTAA AAGTGGACAA CGGTCTGATC GTCGAACTGG 7200  
 ACGGCAAACG CCGGGACCAG TTTGACATGA TCGACCGATT TATCGCCGAT TACGCGATCA 7260  
 ACGTTGAGCG CACAGAGCAG GCAATGCGCC TGGAGGCGGT GGAAATAGCC CGTATGCTGG 7320  
 TGGATATTCA CGTCAGCCGG GAGGAGATCA TTGCCATCAC TACCGCCATC ACGCCGGCCA 7380  
 AAGCGGTCGA GGTGATGGCG CAGATGAACG TGGTGGAGAT GATGATGGCG CTGCAGAAGA 7440  
 TGCGTGCCCG CCGGACCCCC TCCAACCAGT GCCACGTCAC CAATCTCAA GATAATCCGG 7500  
 TGCAGATTGC CGCTGACGCC GCCGAGCCG GGATCCGCGG CTTCTCAGAA CAGGAGACCA 7560  
 CGGTCGGTAT CGCGCGCTAC GCGCGTTTA ACGCCCTGGC GCTGTTGGTC GGTTCGCAGT 7620  
 GCGGCCGCC CCGCGTGTG ACGCAGTGCT CGGTGGAAGA GGCCACCGAG CTGGAGCTGG 7680  
 GCATGCGTGG CTTAACCAGC TACGCCGAGA CGGTGTCGGT CTACGGCACC GAAGCGGTAT 7740

TTACCGACGG CGATGATACG CCGTGGTCAA AGGCGTTCCT CGCCTCGGCC TACGCCTCCC	7800
GCGGGTTGAA AATGCCTAC ACCTCCGGCA CCGGATCCGA AGCGCTGATG GGCTATTCCG	7860
AGAGCAAGTC GATGCTCTAC CTCGAATCGC GCTGCATCTT CATTACTAAA GCGCCCGGG	7920
TTCAGGGACT GCAAAACGGC GCGGTGAGCT GTATCGGCAT GACCGGCGCT GTGCCSTCG	7980
GCATTCGGGC GGTGCTGGCG GAAAACCTGA TCGCCTCTAT GCTCGACCTC GAAGTGGCGT	8040
CCGCCAACGA CCAGACTTTC TCCACTCGG ATATTCCGGC CACCGCGCGC ACCCTGATGC	8100
AGATGCTGCC GGGCACCGAC TTTATTTCT CCGGCTACAG CGCGGTGCCG AACTACGACA	8160
ACATGTTCCG CGGCTCGAAC TTCGATCGCG AAGATTTTGA TGATTACAAC ATCCTGCAGC	8220
GTGACCTGAT GGTGACGGC GGCCTGCGTC CGGTGACCGA GGCGGAAACC ATTGCCATTC	8280
GCCGAAAGC GCGCGGGCG ATCCAGGCGG TTTCCGCGA GCTGGGGCTG CCGCCAATCG	8340
CCGACGAGGA GGTGGAGGCC GCCACCTACG CGCACGGCAG CAACGAGATG CCGCCGCGTA	8400
ACGTGGTGGA GGATCTGAGT GCGGTGGAAG AGATGATGAA GCGCAACATC ACCGGCCTCG	8460
ATATTGTCGG CCGCTGAGC CGCAGCGGCT TTGAGGATAT CGCCAGCAAT ATTCTCAATA	8520
TGCTGCGCCA GCGGGTCACC GCGGATTACC TGCAGACCTC GGCCATTCTC GATCGGCAGT	8580
TCGAGGTGGT GAGTGCCTC AACGACATCA ATGACTATCA GGGCCCGGC ACCGGCTATC	8640
GCATCTCTGC CGAACGCTGG GCGGAGATCA AAAATATTCC GGGCGTGGTT CAGCCCGACA	8700
CCATTGAATA AGGCGGTATT CCTGTGCAAC AGACAACCCA AATTCAGCCC TCTTTTACCC	8760
TGAAAACCCG CGAGGGCGGG GTAGCTTCTG CCGATGAACG CGCCGATGAA GTGGTGATCG	8820
GCGTCGGCCC TGCCTTCGAT AAACACCAGC ATCACACTCT GATCGATATG CCCCATGGCG	8880
CGATCCTCAA AGAGCTGATT GCCGGGGTGG AAGAAGAGGG GCTTCACGCC CGGGTGGTGC	8940
GCATTCTGCG CACGTCCGAC GTCTCCTTTA TGGCCTGGGA TGCGGCCAAC CTGAGCGGCT	9000
CGGGGATCGG CATCGGTATC CAGTCGAAGG GGACCACGGT CATCCATCAG CGCGATCTGC	9060
TGCCGCTCAG CAACCTGGAG CTGTTCTCCC AGGCGCGCT GCTGACGCTG GAGACCTACC	9120
GGCAGATTGG CAAAACGCT GCGCGCTATG CGCGCAAAGA GTCACCTTCG CCGGTGCCGG	9180
TGGTGAACGA TCAGATGGTG CGGCCGAAAT TTATGGCCAA AGCCCGGCTA TTTCATATCA	9240
AAGAGACCAA ACATGTGGTG CAGGACGCCC AGCCCGTCAC CCTGCACATC GACTTAGTAA	9300
GGGAGTGACC ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG	9360
CCCGGAGCAT ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT	9420
GCTCTCTGCG GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA	9480

GGCGCAGATT GCCGAGCAGA TGCAGCGCCA TCGGTTGGCG CGCAATTTCC GCCGCGCGGC 9540  
GGAGCTTATC GCCATTCTCTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TGCGCCCGTT 9600  
CCGCTCCTCG CAGGCGGAGC TGCTGGCGAT CCGCCACGAG CTGGAGCACA CCTGGCATGC 9660  
GACAGTGAAT GCCGCCTTTG TCCGGGAGTC GCGGGAAGTG TATCAGCAGC GGCATAAGCT 9720  
GCGTAAAGGA AGCTAAGCGG AGGTCAGCAT GCCGTTAATA GCCGGGATTG ATATCGGCAA 9780  
CGCCACCACC GAGGTGGCGC TGGCGTCCGA CTACCCGCGC GCGAGGGCGT TTGTTGCCAG 9840  
CGGGATCGTC GCGACGACGG GCATGAAAGG GACGCGGGAC AATATCGCCG GGACCCTCGC 9900  
CGCGCTGGAG CAGGCCCTGG CGAAAACACC GTGGTTCGATG AGCGATGTCT CTCGCATCTA 9960  
TCTTAACGAA GCCGCGCCCG TGATTGGCGA TGTGGCGATG GAGACCATCA CCGAGACCAT 10020  
TATCACCGAA TCGACCATGA TCGGTCATAA CCGCAGACG CCGGGCGGGG TGGGCGTTGG 10080  
CGTGGGGACG ACTATCGCCC TCGGGCGGCT GCGACGCTG CCGGCGGCGC AGTATGCCGA 10140  
GGGGTGGATC GTACTGATTG ACGACGCCGT CGATTTCTT GACGCCGTGT GGTGGCTCAA 10200  
TGAGGCGCTC GACCGGGGGA TCAACGTGGT GCGGCGATC CTCAAAAAGG ACGACGGCGT 10260  
GCTGGTGAAC AACCGCCTGC GTAAAACCTT GCCGGTGGTG GATGAAGTGA CGCTGCTGGA 10320  
GCAGGTCCCC GAGGGGGTAA TGGCGGCGGT GGAAGTGGCC GCGCCGGGCC AGGTGGTGCG 10380  
GATCCTGTCTG AATCCCTACG GGATCGCCAC CTTCTTCGGG CTAAGCCCGG AAGAGACCCA 10440  
GGCCATCGTC CCCATCGCCC GCGCCCTGAT TGGCAACCGT TCCGCGGTGG TGCTCAAGAC 10500  
CCCGCAGGGG GATGTGCAGT CCGGGGTGAT CCGGCGGGC AACCTCTACA TTAGCGGCGA 10560  
AAAGCGCCGC GGAGAGGCCG ATGTCCCGA GGGCGCGGAA GCCATCATGC AGGCGATGAG 10620  
CGCTGCGCT CCGGTACCGG ACATCCCGG CGAACCGGC ACCCACGCCG GCGGCATGCT 10680  
TGAGCGGGTG CGCAAGGTAA TGGCGTCCCT GACCGGCCAT GAGATGAGCG CGATATACAT 10740  
CCAGGATCTG CTGGCGGTGG ATACGTTTAT TCCGCGCAAG GTGCAGGGCG GGATGGCCGG 10800  
CGAGTGCGCC ATGGAGAATG CCGTCGGGAT GCGGCGATG GTGAAAGCGG ATCGTCTGCA 10860  
AATGCAGGTT ATCGCCCGCG AACTGAGCGC CCGACTGCAG ACCGAGGTGG TGGTGGGCGG 10920  
CGTGGAGGCC AACATGGCCA TCGCCGGGGC GTTAACCACT CCCGGCTGTG CGGCGCCGCT 10980  
GGCGATCCTC GACCTCGGCG CCGGCTCGAC GGATGCGGCG ATCGTCAACG CGGAGGGGCA 11040  
GATAACGGCG GTCCATCTCG CCGGGGCGGG GAATATGGTC AGCCTGTTGA TTAAAACCGA 11100  
GCTGGGCCTC GAGGATCTTT CGCTGGCGGA AGCGATAAAA AAATACCCGC TGGCCAAAGT 11160  
GGAAAGCCTG TTCAGTATTC GTCACGAGAA TGGCGCGGTG GAGTCTTTC GGAAGCCCT 11220

```

CAGCCCGGCG GTGTTGCGCA AAGTGGTGTA CATCAAGGAG GCGGAACTGG TGCCGATCGA 11280
TAACGCCAGC CCGCTGGAAA AAATTCGTCT CGTGCGCCGG CAGGCGAAAAG AGAAAGTGTT 11340
TGTCACCAAC TGCCTGCGCG CGCTGCGCCA GGTCTCACCC GCGGGTTCCA TTCGCGATAT 11400
CGCCTTTGTG GTGCTGGTGG GCGGCTCATC GCTGGACTTT GAGATCCCGC AGCTTATCAC 11460
GGAAGCCTTG TCGACTATG GCGTGGTCGC CGGGCAGGGC AATATTCGGG GAACAGAAGG 11520
GCCGCGCAAT GCGGTCGCCA CCGGGCTGCT ACTGGCCGGT CAGGCGAATT AAACGGGCGC 11580
TCGCGCCAGC CTCTCTCTTT AACGTGCTAT TTCAGGATGC CGATAATGAA CCAGACTTCT 11640
ACCTTAACCG GGCAGTGCCT GGCCGAGTTT CTTGGCACCG GATTGCTCAT TTTCTTCGGC 11700
GCGGGCTGCG TCGCTGCGCT GCGGGTCGCC GGGGCCAGCT TTGGTCAGTG GGAGATCAGT 11760
ATTATCTGGG GCCTTGGCGT CGCCATGGCC ATCTACCTGA CGGCCGGTGT CTCCGGCGCG 11820
CACCTAAATC CGGCGGTGAC CATTGCCCTG TGGCTGTTTC CCTGTTTTGA ACGCCGCAAG 11880
GTGCTGCCGT TTATTGTTGC CCAGACGGCC GGGGCCTTCT GCGCCGCCGC GCTGGTGTAT 11940
GGGCTCTATC GCCAGCTGTT TCTCGATCTT GAACAGAGTC AGCATATCGT GCGCGGCACT 12000
GCCGCCAGTC TTAACCTGGC CGGGGTCTTT TCCACGTACC CGCATCCACA TATCACTTTT 12060
ATACAAGCGT TTGCCGTGGA GACCACCATC ACGGCAATCC TGATGGCGAT GATCATGGCC 12120
CTGACCGACG ACGGCAACGG AATTC 12145

```

## (2) SEQ ID NO: 11 的信息

### (i) 序列特征

(A) 长度: 94 碱基对

(B) 类型: 核酸

(C) 链: 双链

(D) 拓扑性质: 线性

(ii) 分子类型: DNA (基因组)

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 11

```

AGCTTAGGAG TCTAGAATAT TGAGCTCGAA TTCCCGGGCA TGCGGTACCG GATCCAGAAA 60
AAAGCCCGCA CCTGACAGTG CCGGCTTTTT TTTT 94

```

## (2) SEQ ID NO: 12 的信息

## (i) 序列特征

- (A) 长度: 37 碱基对
  - (B) 类型: 核酸
  - (C) 链: 单链
  - (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
- (A) 描述: /desc="引物'
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 12

GGAATTCAGA TCTCAGCAAT GAGCGAGAAA ACCATGC

37

## (2) SEQ ID NO: 13 的信息

## (i) 序列特征

- (A) 长度: 27 碱基对
  - (B) 类型: 核酸
  - (C) 链: 单链
  - (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
- (A) 描述: /desc="引物'
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 13

GCTCTAGATT AGCTTCCTTT ACGCAGC

27

## (2) SEQ ID NO: 14 的信息

## (i) 序列特征

- (A) 长度: 33 碱基对
- (B) 类型: 核酸

- (C) 链: 单链
- (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
  - (A) 描述: /desc="引物'
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 14

GGCCAAGCTT AAGGAGGTTA ATTAAATGAA AAG

33

## (2) SEQ ID NO: 15 的信息

- (i) 序列特征
  - (A) 长度: 26 碱基对
  - (B) 类型: 核酸
  - (C) 链: 单链
  - (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
  - (A) 描述: /desc="引物'
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 15

GCTCTAGATT ATTCAATGGT GTCGGG

26

## (2) SEQ ID NO: 16 的信息

- (i) 序列特征
  - (A) 长度: 42 碱基对
  - (B) 类型: 核酸
  - (C) 链: 单链
  - (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
  - (A) 描述: /desc="引物'

- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 16

GCGCCGTCTA GAATTATGAG CTATCGTATG TTTGATTATC TG

42

## (2) SEQ ID NO: 17 的信息

- (i) 序列特征
  - (A) 长度: 36 碱基对
  - (B) 类型: 核酸
  - (C) 链: 单链
  - (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
  - (A) 描述: /desc="引物"
- (iii) 是否假设: 否
- (vi) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 17

TCTGATACGG GATCCTCAGA ATGCCTGGCG GAAAAT

36

## (2) SEQ ID NO: 18 的信息

- (i) 序列特征
  - (A) 长度: 30 碱基对
  - (B) 类型: 核酸
  - (C) 链: 单链
  - (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
  - (A) 描述: /desc="引物"
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 18

TCTATTGTGG ATGCTTTACC ATGGTTAAAA

30

(2) SEQ ID NO: 19 的信息

(i) 序列特征

(A) 长度: 26 碱基对

(B) 类型: 核酸

(C) 链: 单链

(D) 拓扑性质: 线性

(ii) 分子类型: 其它核酸

(A) 描述: /desc="引物'

(v) 是否假设: 否

(vi) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 19

CACCGACGCC GGATCCAAAC ACCAGC

26

(2) SEQ ID NO: 20 的信息

(i) 序列特征

(A) 长度: 38 碱基对

(B) 类型: 核酸

(C) 链: 单链

(D) 拓扑性质: 线性

(ii) 分子类型: 其它核酸

(A) 描述: /desc="引物'

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 20

TCACTGTCCA AGAGGATCCG TAAAATCAAC GCCATGAC

38

## (2) SEQ ID NO: 21 的信息

## (i) 序列特征

(A) 长度: 34 碱基对

(B) 类型: 核酸

(C) 链: 单链

(D) 拓扑性质: 线性

## (ii) 分子类型: 其它核酸

(A) 描述: /desc="引物'

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 21

GGCATTGGC GGCGAAGCTT TATGGTGGCT ACAC

34

## (2) SEQ ID NO: 22 的信息

## (i) 序列特征

(A) 长度: 18 碱基对

(B) 类型: 核酸

(C) 链: 单链

(D) 拓扑性质: 线性

## (ii) 分子类型: 其它核酸

(A) 描述: /desc="引物'

(iii) 是否假设: 否

(iv) 是否反义: 否

(xi) 序列描述: SEQ ID NO: 22

TCGACGAATT CAGGAGGA

18

## (2) SEQ ID NO: 23 的信息

## (i) 序列特征

(A) 长度: 18 碱基对

(B) 类型: 核酸

- (C) 链: 单链
- (D) 拓扑性质: 线性
- (ii) 分子类型: 其它核酸
  - (A) 描述: /desc="引物'
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 23

CTAGTCCTCC TGAATTCG

18

## (2) SEQ ID NO: 24 的信息

- (i) 序列特征
  - (A) 长度: 4549 碱基对
  - (B) 类型: 核酸
  - (C) 链: 双链
  - (D) 拓扑性质: 线性
- (ii) 分子类型: DNA (基因组)
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 24

AGCTCGTCAG CGGGTGTGG CGGGTGTGG GGCTGGCTTA ACTATGCGGC ATCAGAGCAG 60  
ATTGTA CTGA GAGTGCACCA TATGCGGTGT GAAATACCGC ACAGATGCGT AAGGAGAAAA 120  
TACCGCATCA GCGGCCATTC GCCATT CAGG CTGCGCAACT GTTGGGAAGG GCGATCGGTG 180  
CGGGCCTCTT CGCTATTACG CCAGCTGGCG AAAGGGGGAT GTGCTGCAAG GCGATTAAGT 240  
TGGGTAACGC CAGGGTTTTTC CCAGTCACGA CGTTGTAAAA CGACGGCCAG TGAATTGCGAG 300  
CTCGGTACCC GGGGATCCTC TAGAGTCGAC CTGCAGGCAT GCAAGCTTGG CGTAATCATG 360  
GTCATAGCTG TTTCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC 420  
CGGAAGCATA AAGTGTAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC 480  
GTTGCGCTCA CTGCCGCTT TCCAGTCGGG AAACCTGTCTG TGCCAGCTGC ATTAATGAAT 540  
CGGCCAACGC GAATTCCCGA CAGTAAGACG GGTAAGCCTG TTGATGATAC CGCTGCCTTA 600  
CTGGGTGCAT TAGCCAGTCT GAATGACCTG TCACGGGATA ATCCGAAGTG GTCAGACTGG 660  
AAAATCAGAG GGCAGGAAGT GCTGAACAGC AAAAAGTCAG ATAGCACCAC ATAGCAGACC 720  
CGCCATAAAA CGCCCTGAGA AGCCCGTGAC GGGCTTTTCT TGTATTATGG GTAGTTTCCT 780  
TGCATGAATC CATAAAAGGC GCCTGTAGTG CCATTTACCC CCATTCCTG CCAGAGCCGT 840  
GAGCGCAGCG AACTGAATGT CACGAAAAAG ACAGCGACTC AGGTGCCTGA TGGTCGGAGA 900  
CAAAAGGAAT ATTCAGCGAT TTGCCGAGC TTGCGAGGGT GCTACTTAAG CCTTTAGGGT 960  
TTTAAGGTCT GTTTTGTAGA GGAGCAAACA GCGTTTGCGA CATCCTTTTG TAATACTGCG 1020  
GAACTGACTA AAGTAGTGAG TTATACACAG GGCTGGGATC TATTCTTTTT ATCTTTTTTT 1080  
ATTCTTTCTT TATTCTATAA ATTATAACCA CTTGAATATA AACAAAAAAA ACACACAAAG 1140  
GTCTAGCGGA ATTTACAGAG GGTCTAGCAG AATTTACAAG TTTTCAGCA AAGGTCTAGC 1200  
AGAATTTACA GATACCCACA ACTCAAAGGA AAAGGACTAG TAATTATCAT TGA CTAGCCC 1260  
ATCTCAATTG GTATAGTGAT TAAAATCACC TAGACCAATT GAGATGTATG TCTGAATTAG 1320  
TTGTTTTCAA AGCAAATGAA CTAGCGATTA GTCGCTATGA CTTAACGGAG CATGAAACCA 1380  
AGCTAATTTT ATGCTGTGTG GCACTACTCA ACCCCACGAT TGAAAACCTT ACAAGGAAAG 1440

AACGGACGGT ATCGTTCCT TATAACCAAT ACGCTCAGAT GATGAACATC AGTAGGGAAA 1500  
ATGCTTATGG TGTATTAGCT AAAGCAACCA GAGAGCTGAT GACGAGAACT GTGGAAATCA 1560  
GGAATCCTTT GGTAAAGGC TTGAGATTI TCCAGTGGAC AAACATATGCC AAGTTCTCAA 1620  
GCGAAAAATT AGAATTAGTT TTAGTGAAG AGATATTGCC TTATCTTTTC CAGTTAAAAA 1680  
AATTCATAAA ATATAATCTG GAACATGTTA AGTCTTTTGA AAACAAATAC TCTATGAGGA 1740  
TTTATGAGTG GTTATTAAAA GAACTAACAC AAAAGAAAAC TCACAAGGCA AATATAGAGA 1800  
TTAGCCTTGA TGAATTTAAG TTCATGTTAA TGCTTGAAAA TAACTACCAT GAGTTTAAAA 1860  
GGCTTAACCA ATGGGTTTTG AAACCAATAA GTAAAGATTT AAACACTTAC AGCAATATGA 1920  
AATTGGTGGT TGATAAGCGA GGCCGCCCGA CTGATACGTT GATTTTCCAA GTTGAAGTAG 1980  
ATAGACAAAT GGATCTCGTA ACCGAAGTTG AGAACAACCA GATAAAAATG AATGGTGACA 2040  
AAATACCAAC AACCATTACA TCAGATTCCT ACCTACATAA CGGACTAAGA AAAACACTAC 2100  
ACGATGCTTT AACTGCAAAA ATTCAGCTCA CCAGTTTGA GGCAAAATTT TTGAGTGACA 2160  
TGCAAAGTAA GTATGATCTC AATGGTTCGT TCTCATGGCT CACGCAAAAA CAACGAACCA 2220  
CACTAGAGAA CATACTGGCT AAATACGGAA GGATCTGAGG TTCTTATGGC TCTTGTATCT 2280  
ATCAGTGAAG CATCAAGACT AACAAACAAA AGTAGAACAA CTGTTACCCG TTACATATCA 2340  
AAGGGAAAAC TGCCCATATG CACAGATGAA AACGGTGTA AAAAGATAGA TACATCAGAG 2400  
CTTTTACGAG TTTTGGTGC ATTCAAAGCT GTTCACCATG AACAGATCGA CAATGTAAACA 2460  
GATGAACAGC ATGTAACACC TAATAGAACA GGTGAAACCA GTAAAACAAA GCAACTAGAA 2520  
CATGAAATTG AACACCTGAG ACAACTTGTT ACAGCTCAAC AGTCACACAT AGACAGCCTG 2580  
AAACAGCGGA TGCTGCTTAT CGAATCAAAG CTGCCGACAA CACGGGAGCC AGTGACGCCT 2640  
CCCGTGGGGA AAAAATCATG GCAATTCTGG AAGAAATAGC GCTTTCAGCC GGCAAACCGG 2700  
CTGAAGCCGG ATCTGCGATT CTGATAACAA ACTAGCAACA CCAGAACAGC CCGTTTCCGG 2760  
GCAGCAAAAC CCGTGGGAAT TAATCCCCT GCTCGGCAG GCTGGGTGCC AAGCTCTCGG 2820  
GTAACATCAA GGCCCGATCC TTGGAGCCCT TGCCCTCCCG CACGATGATC GTGCCGTGAT 2880  
CGAAATCCAG ATCCTTGACC CGCAGTTGCA AACCTCACT GATCCGCATG CCCGTTCCAT 2940  
ACAGAAGCTG GGCGAACAAA CGATGCTCGC CTTCAGAAA ACCGAGGATG CGAACCCTT 3000  
CATCCGGGGT CAGCACCACC GGCAAGCGCC GCGACGGCCG AGGTCTTCCG ATCTCCTGAA 3060  
GCCAGGGCAG ATCCGTGCAC AGCACCTTGC CGTAGAAGAA CAGCAAGGCC GCCAATGCCT 3120  
GACGATGCGT GGAGACCGAA ACCTTGCCT CGTTCGCCAG CCAGGACAGA AATGCCTCGA 3180

```

CTTCGCTGCT GCCCAAGGTT GCCGGGTGAC GCACACCGTG GAAACGGATG AAGGCACGAA 3240
CCCAGTGGAC ATAAGCCTGT TCGGTTTCGTA AGCTGTAATG CAAGTAGCGT ATGCGCTCAC 3300
GCAACTGGTC CAGAACCTTG ACCGAACGCA GCGGTGGTAA CGGCGCAGTG GCGGTTTTCA 3360
TGGCTTGTTA TGA CTGTTTT TTTGGGGTAC AGTCTATGCC TCGGGCATCC AAGCAGCAAG 3420
CGCGTTACGC CGTGGGTCGA TGTTTGATGT TATGGAGCAG CAACGATGTT ACGCAGCAGG 3480
GCAGTCGCCC TAAACAAAG TTAACATCA TGAGGGAAGC GGTGATCGCC GAAGTATCGA 3540
CTCAACTATC AGAGGTAGTT GGCATCATCG AGCGCCATCT CGAACCGACG TTGCTGGCCG 3600
TACATTTGTA CGGCTCCGCA GTGGATGGCG GCCTGAAGCC ACACAGTGAT ATTGATTTGC 3660
TGGTTACGGT GACCGTAAGG CTTGATGAAA CAACGCGGCG AGCTTTGATC AACGACCTTT 3720
TGGAAACTTC GGCTTCCCCT GGAGAGAGCG AGATTCTCCG CGCTGTAGAA GTCACCATTG 3780
TTGTGCACGA CGACATCATT CCGTGGCGTT ATCCAGCTAA GCGCGAACTG CAATTTGGAG 3840
AATGGCAGCG CAATGACATT CTTGCAGGTA TCTTCGAGCC AGCCACGATC GACATTGATC 3900
TGGCTATCTT GCTGACAAAA GCAAGAGAAC ATAGCGTTGC CTTGGTAGGT CCAGCGGCGG 3960
AGGAACTCTT TGATCCGGTT CCTGAACAGG ATCTATTTGA GCGGCTAAAT GAAACCTTAA 4020
CGCTATGGAA CTCGCCGCC GACTGGGCTG GCGATGAGCG AAATGTAGTG CTTACGTTGT 4080
CCCGCATTG GTACAGCGCA GTAACCGCA AAATCGCGCC GAAGGATGTC GCTGCCGACT 4140
GGGCAATGGA GCGCCTGCCG GCCCAGTATC AGCCCATCAT ACTTGAAGCT AGACAGGCTT 4200
ATCTTGGACA AGAAGAAGAT CGCTTGGCCT CGCGCGCAGA TCAGTTGGAA GAATTTGTCC 4260
ACTACGTGAA AGGCGAGATC ACCAAGGTAG TCGGCAAATA ATGTCTAACA ATTCGTTCAA 4320
GCCGACGCCG CTTCGCGGCG CGGCTTAACT CAAGCGTTAG ATGCACTAAG CACATAATTG 4380
CTCACAGCCA AACTATCAGG TCAAGTCTGC TTTTATTATT TTTAAGCGTG CATAATAAGC 4440
CCTACACAAA TTGGGAGATA TATCATGAAA GGCTGGCTTT TTCTTGTTAT CGCAATAGTT 4500
GGCGAAGTAA TCGCAACATC CGCATTAAAA TCTAGCGAGG GCTTTACTA 4549

```

## (2) SEQ ID NO: 25 的信息

### (i) 序列特征

- (A) 长度: 199 碱基对
- (B) 类型: 核酸
- (C) 链: 双链
- (D) 拓扑性质: 线性

- (ii) 分子类型: DNA (基因组)
- (iii) 是否假设: 否
- (iv) 是否反义: 否
- (xi) 序列描述: SEQ ID NO: 25

```
GAATTCACTA GTCGATCTGT GCTGTTTGCC ACGGTATGCA GCACCAGCGC GAGATTATGG 60
GCTCGCACGC TCGACTGTCTG GACGGGGGCA CTGGAACGAG AAGTCAGGCG AGCCGTCACG 120
CCCTTGACAA TGCCACATCC TGAGCAAATA ATTCAACCAC TAAACAAATC AACCGCGTTT 180
CCCGGAGGTA ACCAAGCTT 199
```