

[19]中华人民共和国专利局

[51]Int.Cl⁶



[12] 发明专利申请公开说明书

[21] 申请号 97109799.2

C12N 15/60
C12N 15/74 C12N 9/88
C12N 1/21 C12P 17/18
C12P 17/06 A23K 1/17

[43]公开日 1998 年 1 月 28 日

[11] 公开号 CN 1171443A

[22]申请日 97.5.6

[30]优先权

[32]96.5.7 [33]US[31]60/016,753

[32]97.4.4 [33]US[31]042,935

[71]申请人 弗·哈夫曼-拉罗切有限公司

地址 瑞士巴塞尔

[72]发明人 C·D·里夫斯 C·L·索丽黛

[74]专利代理机构 中国国际贸易促进委员会专利商标
事务所
代理人 李 瑛

权利要求书 2 页 说明书 79 页 附图页数 5 页

[54]发明名称 富伦霉素基因簇

[57]摘要

本发明提供了含有编码酶的 DNA 序列, 所述酶在原核细胞中催化富伦霉素的生物合成, 提供了含有所述 DNA 序列的载体, 由所述 DNA 序列编码的蛋白质, 由所述 DNA 序列转化的宿主细胞以及利用所述转化细胞生产富伦霉素, 尤其是富伦霉素 B 的方法。

权 利 要 求 书

1.一种 DNA 序列, 含有至少下列 DNA 序列之一:

- a) SEQ ID NO: 22 的 636 到 2948 的碱基,
- b) SEQ ID NO: 22 的 2945 到 3916 的碱基,
- c) SEQ ID NO: 22 的 4020 到 4844 的碱基,
- d) SEQ ID NO: 22 的 4841 到 6415 的碱基,
- e) SEQ ID NO: 22 的 6533 到 7183 的碱基,
- f) SEQ ID NO: 22 的 7344 到 8897 的碱基,
- g) SEQ ID NO: 22 的 9164 到 10012 的碱基,
- h) SEQ ID NO: 22 的 10621 到 10105 的碱基,
- i) SEQ ID NO: 22 的 11628 到 10618 的碱基,
- k) SEQ ID NO: 22 的 11809 到 12066 的碱基,
- l) SEQ ID NO: 22 的 13209 到 12154 的碱基,
- m) SEQ ID NO: 22 的 13409 到 14686 的碱基,
- n) SEQ ID NO: 22 的 14767 到 16047 的碱基,
- o) SEQ ID NO: 22 的 16120 到 16371 的碱基,
- p) SEQ ID NO: 22 的 16935 到 16453 的碱基,
- q) SEQ ID NO: 22 的 17088 到 17903 的碱基,
- r) SEQ ID NO: 22 的 17903 到 18898 的碱基,
- s) SEQ ID NO: 22 的 18895 到 19839 的碱基,
- t) SEQ ID NO: 22 的 20907 到 19990 的碱基,
- w) SEQ ID NO: 22 的 22094 到 20904 的碱基,

或者含有一种 DNA 序列, 它编码由 a)-w)限定的 DNA 序列编码的蛋白质的生化特性等同的变异体。

2.根据权利要求 1 所述的 DNA 序列, 其中含有由 a)-w)限定的所有 DNA 序列, 优选的是他们的结构类似于 SEQ ID NO: 22 的描述。或者含有这样的 DNA, 其中至少一种由 a)-w)限定的 DNA 序列被一种 DNA 序列替代, 用于替代的 DNA 序列编码一种其生化特性等同变异体的蛋白质。

3. 根据权利要求 2 所述的 DNA 序列, 其中含有 SEQ ID NO: 22 的 DNA 序列或者一种编码其生化特性等同变异体的蛋白质的 DNA。

4. 含有权利要求 1 - 3 中任何一项所述的 DNA 序列的载体, 其中所述的 DNA 与表达控制序列可操作地连接。

5. 用权利要求 4 定义的载体转化的宿主细胞。

6. 权利要求 5 所述的转化细胞, 其中宿主细胞是链霉菌属的成员。

7. 权利要求 6 所述的转化细胞, 其中宿主细胞是玫瑰暗黄色链霉菌。

8. 一种蛋白质和其生化特性功能等同变异体, 它由权利要求 1 中 a)-w) 限定的 DNA 序列编码或者具有 SEQ ID NO: 1 - 21 给出的氨基酸序列。

9. 用于制备富伦霉素或所述富伦霉素的生物合成中间体的方法, 其特征在于在合适的培养条件下培养权利要求 5 - 7 中任何一项所述的细胞并且从培养物或细胞分离富伦霉素。

10. 用于制备富伦霉素的方法, 其特征在于由权利要求 9 所述方法获得的生物合成中间体用本领域内技术人员已知的化学方法转化为所述富伦霉素。

11. 用于制备富伦霉素 B 的方法, 其中将权利要求 9 或 10 所述方法获得的或富伦霉素混合物氧化为富伦霉素 B。

12. 用于制备饲料组合物的方法, 其特征在于实施权利要求 9 - 11 所述方法并且将获得的富伦霉素与其他饲料组合物配料混合。

13. 本文中前面描述的发明。

说明书

富伦霉素基因簇

本发明涉及与微生物方法生产的抗菌素富伦霉素 B，特别是从玫瑰暗黄链霉菌生产的抗菌素富伦霉素 B 相关的基因的鉴定，分离和测序。

聚酮化合物 是一个大的，结构上有各种变化的天然产物的家族，显示具有各种各样的生物学活性包括抗菌性和药理学特性。例如，聚酮化合物的典型代表是抗菌素四环素和红霉素，抗癌剂柔毛霉素，免疫抑制剂 FK506 和纳巴霉素，和兽药制品例如富伦霉素，莫能菌素和阿夫曼菌素。

聚酮化合物合成酶(PKSs)是与脂肪酸合成酶(FASs)相关的多功能酶。PKSs 通过使酰基硫酸酯，通常是乙酰基，丙酰基，丙二酰基或甲基丙二酰基之间重复进行 claisen 缩合而催化聚酮化合物的生物合成。在执行每个缩合步骤时，他们通过催化整个还原环或其部分或没有催化还原环包括在逐渐形成的聚酮化合物链的 β -酮基进行酮基还原，脱水，和烯酰基还原而在该产物中导入了结构变异性。在碳链形成了各个特定产物的长度特征后，通过硫解和酰基转移而从合成酶中释放出碳链。因此，PKSs 由在一起工作而产生给定聚酮化合物的酶家族组成。在链长，构链单位的选择，以及按遗传计划将还原环进入各个 PKS 的受控制的变异归功于在天然存在的聚酮化合物之间看到的变异。

链霉菌属是芳香族聚酮化合物的放线菌生产者；在链霉菌属中，已知 PKS 介导的合成依赖于至少三个 PKS 开放式阅读框架(ORFs)的产物。PRF1 编码酮基合成酶(KS)和酰基转移酶(AT)的活性位点；ORF2 编码类似于 ORF1 产物但失去 KS 和 AT 基序的蛋白质；ORF3 编码一个分离的酰基载体蛋白(ACP)。

富伦霉素 B 是广泛地用作为家养动物和鸟类的抗球孢子药的萘醌聚酮化合物。在 J.Antibiotics,31:10,959-965(1978)中 Iwai 等人报道了从玫瑰暗黄色链霉菌菌株 No.AM-3867，一种土壤分离株的发酵肉汤中分离富伦霉素组的两种抗菌素 AM-3867 I 和 II。AM-3867 I 被发现是一种后来命名为富伦霉素 B 的新抗菌素，而后者 AM-3867 II 被鉴定为脱氧富伦霉素。

Bibb 等人在 *Gene*,142,31-39(1994)报道了来自玫瑰暗黄色链霉菌的 10.2kb 的 DNA 片段的克隆, 据推测该片段含有编码聚酮化合物合成酶 (PKS) 的基因(fren), 该基因决定富伦霉素和七尾霉素的产生。该 DNA 的一个 5530bp 的连续片段被测序了。对该序列分析结果显示五个完整的开放式阅读框架(ORFs)以一个方向转录(ORFs 1,2,3,5,4)以及位于 ORF3 和 ORF5 之间的一个(ORFX)以另一相反方向转录。推测到的 ORFs 1, 2, 3, 4 和 5 的氨基酸序列分别类似接近于来自链霉菌属的 II 型 PKSs 的已知成分的序列: 假设的异源二聚体(ORF1+2)酮基合成酶, 酰基载体蛋白, 环化酶和酮还原酶。ORF4 产物的 N - 末端和 C - 末端一半之间的相似性也发现于来自其他异氧杂萜满苯醌抗菌素生产者的相应基因中, 这暗示着编码环化酶的基因可能有成双源点。ORFX 似乎代表不仅存在于 fren 簇, 而且还存在于链霉菌各个种的其他簇的芳香族抗菌素生物合成基因中的未知功能的新的一类基因。据报道通过基因分裂的方法将这些基因导入到 AK24158 菌株中以便证明 fren 基因的确是那些编码富伦霉素 PKS 的基因的尝试没有成功。

本发明的一个目的是提供编码富伦霉素的基因簇的全长序列并且分离和鉴定了该基因簇的开放式阅读框架。

本发明的另一目的是提供用任意或所有分离到的富伦霉素 PKS 基因转化的遗传工程方法得到的宿主细胞。

本发明的再一目的是改善富伦霉素生物合成生产的产量。

本发明提供了一个分离到的基因簇, 该基因簇包括编码在原核细胞中催化富伦霉素生物合成的蛋白质的 SEQ ID NO: 22, 并且提供了经修饰的基因簇, 该基因簇与 SEQ ID NO: 22 具有足够的生化功能等同性, 另外它也编码催化富伦霉素生物合成的蛋白质。还包括含有可操作连接到表达控制序列上的基因簇的载体以及由该载体转化的宿主细胞。

还提供了五个分离出的基因亚簇, 他们负责富伦霉素合成的特定方面:

编码用于生产富伦霉素的溢出泵的离体的第一基因亚簇, 它包括编码列于 SEQ ID NOS: 1, 2, 3, 4, 和 6 中的蛋白质和所说蛋白质的生化特性等同突变蛋白的基因;

编码用于生产富伦霉素的丁酸起始物合成酶的离体的第二基因亚簇，它包括编码列于 SEQ ID NOS: 8, 9, 10, 和 11 中的蛋白质和所说蛋白质的生化特性等同突变蛋白的基因；

编码用于生产富伦霉素的聚酮化合物合成酶的离体的第三基因亚簇，它包括编码列于 SEQ ID NOS: 12, 13 和 14 中的蛋白质和所说蛋白质的生化特性等同突变蛋白的基因；

编码用于生产富伦霉素的半缩酮酶，酮基还原酶，和环化酶/脱氢酶的离体的第四基因亚簇，它包括编码列于 SEQ ID NOS: 15, 16, 17, 和 18 中的蛋白质和所说蛋白质的生化特性等同突变蛋白的基因；

编码用于生产富伦霉素的酮基/烯酰还原酶和羟化酶的离体的第五基因亚簇，它包括编码列于 SEQ ID NOS: 19 和 20 中的蛋白质和所说蛋白质的生化特性等同突变蛋白的基因。

在另一个实施方案中，将各个基因单独地插入到载体中并且可操作地连接到表达控制序列上。

在另一个实施方案中，提供了富伦霉素的生物合成的方法，该方法包括 a) 用含有 SEQ ID NO: 22 基因簇的重组载体转化宿主细胞； b) 在所说宿主细胞中表达所说载体；和 c) 分离由此产生的富伦霉素。

还提供了重组生物合成抗菌素富伦霉素 B 的方法，该方法包括 a) 用含有 SEQ ID NO: 22 基因簇的载体转化宿主细胞； b) 在所说宿主细胞中表达所说载体；和 c) 分离由此产生的富伦霉素； d) 将富伦霉素混合物氧化以便基本上将所有分离到的产物转化为富伦霉素 B。

还提供了 SEQ ID NOS: 1 - 20 的分离到的蛋白质以及其生化特性等同变异体，提供了编码所说蛋白质的离体基因。

本发明的又一目的是提供了含有至少下列序列之一的 DNA 序列：

- a) SEQ ID NO: 22 的 636 到 2948 的碱基。
- b) SEQ ID NO: 22 的 2945 到 3916 的碱基。
- c) SEQ ID NO: 22 的 4020 到 4844 的碱基。
- d) SEQ ID NO: 22 的 4841 到 6415 的碱基。
- e) SEQ ID NO: 22 的 6533 到 7183 的碱基。
- f) SEQ ID NO: 22 的 7344 到 8897 的碱基。

- g) SEQ ID NO: 22 的 9164 到 10012 的碱基。
- h) SEQ ID NO: 22 的 10621 到 10105 的碱基。
- i) SEQ ID NO: 22 的 11628 到 10618 的碱基。
- k) SEQ ID NO: 22 的 11809 到 12066 的碱基。
- l) SEQ ID NO: 22 的 13209 到 12154 的碱基。
- m) SEQ ID NO: 22 的 13409 到 14686 的碱基。
- n) SEQ ID NO: 22 的 14767 到 16047 的碱基。
- o) SEQ ID NO: 22 的 16120 到 16371 的碱基。
- p) SEQ ID NO: 22 的 16935 到 16453 的碱基。
- q) SEQ ID NO: 22 的 17088 到 17903 的碱基。
- r) SEQ ID NO: 22 的 17903 到 18898 的碱基。
- s) SEQ ID NO: 22 的 18895 到 19839 的碱基。
- t) SEQ ID NO: 22 的 20907 到 19990 的碱基。
- w) SEQ ID NO: 22 的 22094 到 20904 的碱基。

或者提供了为具有与 a)-w)限定的 DNA 序列编码的蛋白质等同的生化特性的变异体编码的 DNA 序列, 更具体地说, 提供了这样一种 DNA 序列, 它包括 a)-w)限定的, 优选地以相同于 SEQ ID NO: 22 的方式描述的所有 DNA 序列, 或者提供了这样一种 DNA 序列, 其中至少由 a)-w)限定的 DNA 序列之一由编码其生化特性等同变异体的蛋白质的 DNA 序列所替代, 进一步具体地说提供了一种 DNA 序列, 它包括 SEQ ID NO: 22 的 DNA 序列, 或者一种 DNA 序列, 它编码是其生化特性等同变异体的蛋白质。

本发明的再一目的是提供了含有可操作连接到表达控制序列上的所说 DNA 序列的载体, 提供了用所说载体转化的宿主细胞, 特别是其中宿主细胞是链霉菌属的成员, 例如玫瑰暗黄色链霉菌。

本发明的还有一个目的是提供了由上述 a)-w)定义的 DNA 序列编码的蛋白质或者具有 SEQ ID NO: 1 - 21 给出的氨基酸序列和其生化特性等同的变异体的蛋白质, 提供了用于制备富伦霉素或所说富伦霉素的生物合成中间体的方法, 该方法的特征在于在合适的培养条件下培养上述定义的细胞以及从所说培养物或细胞中分离富伦霉素, 以及提供了制备富伦霉素的方法, 其特征在于采用本领域内已知的化学和其他方法将上述定义的方法

法获得的生物合成中间体转化为所说的富伦霉素，提供了制备富伦霉素 B 的方法，其中将上文中定义的方法获得的富伦霉素或富伦霉素的混合物氧化为富伦霉素 B。

本发明的还有一个目的是提供了制备饲料组合物的方法，其特征在于实施上文中定义的一种方法，将获得的富伦霉素与其他的饲料组合物配料混合。

下面的附图用于阐明本发明：

附图 1 富伦霉素基因簇的限制性酶切图谱。

附图 2 假设的富伦霉素生物合成途径，其中说明了各种基因亚簇的功能。

附图 3 转化载体 pSSVtsr 和 pSSVaph 的构建。

附图 4 合成的 ermE 和 tipA 启动子的序列。

附图 5 用于在玫瑰暗黄色链霉菌中进行基因表达的载体的构建。这里显示了转录活化基因构建体。

下列定义用于阐明和限定各种术语的含义和范围，所说术语用于描述本发明。

“宿主细胞”是从原核微生物衍生得到的细胞，它可用作为重组载体的受体，所说载体含有 DNA，尤其是本发明富伦霉素基因簇。该术语包括已被转化的原始细胞的子代。由于偶然的或有意的突变，单个亲代细胞的子代在形态上或其基因组或总 DNA 装备不是必然完全等同于原始亲代。该定义也包括完全类似于亲本的亲代细胞的子代，所谓类似是由相关特性例如存在特定的 DNA 序列定性的。

针对核酸序列例如编码序列，控制序列而言时术语“异源的”是指正常情况下不与重组构建体的一个区域相关的序列，和/或正常情况下不与特定细胞相关的序列。因此，一个核酸构建体的“异源的”区域是另一个核酸分子内或吸附到另一个核酸分子上的一个可鉴定的核酸片段，在自然界中它不与其他分子结合存在。例如一种构建体的异源区域包括编码区，在自然状态其两侧序列不与该编码序列结合存在。异源编码序列的另一个例子是其中的编码序列本身不存在于自然界中的构建体(例如，具有不同于天然基因的密码子的合成序列)。类似地，用正常情况下不存在于宿主细胞中

的构建体转化的宿主细胞被认为适用于本发明目的异源。等位变异或自然出现的突变事件并不产生异源 DNA，如本文所述。

“编码序列”是指一种核酸序列，在置于合适的调节序列控制下该核酸序列可在体内或体外转录(对于 DNA)或转译为(对于 mRNA)多肽。由 5' (氨基)末端的转译起始密码子和 3' (羧基)末端转译终止子密码子决定了该编码序列的边界。通常转录终止序列定位于编码序列的 3' 末端。

“核酸序列”包括，但不限于，原核和真核 mRNA，来自原核和真核 mRNA 的 cDNA，基因组 DNA，和合成 DNA 和 RNA 序列，他们含有天然的核苷碱基腺嘌呤，鸟嘌呤，胞嘧啶，胸腺嘧啶和尿嘧啶。该术语还包括含有一个或多个其他碱基的序列，包括但不限于 4 - 乙酰基胞嘧啶，8 - 羟基 - N6 - 甲基腺嘌呤，氮丙啶基胞嘧啶，假异胞嘧啶，5 - (羧基羟基甲基)尿嘧啶，5 - 氟尿嘧啶，5 - 溴尿嘧啶，5 - 羧基羟基甲基氨基甲基 - 2 - 硫尿嘧啶，5 - 羧基甲基氨基甲基尿嘧啶，二氢尿嘧啶，次黄苷，N6 - 异戊烯基 - 腺嘌呤，1 - 甲基腺嘌呤，1 - 甲基假尿嘧啶，1 - 甲基鸟嘌呤，1 - 甲基 - 次黄苷，2，2 - 二甲基鸟嘌呤，2 - 甲基腺嘌呤，2 - 甲基鸟嘌呤，3 - 甲基 - 胞嘧啶，5 - 甲基胞嘧啶，N6 - 甲基腺嘌呤，7 - 甲基鸟嘌呤，5 - 甲基氨基甲基胞嘧啶，5 - 甲氧基氨基甲基 - 2 - 硫尿嘧啶， β - D - 甘露糖 queosine，5' - 甲氧基羧基甲基尿嘧啶，5 - 甲氧基尿嘧啶，2 - 甲基硫 - N6 - 异戊烯基腺嘌呤，尿嘧啶 - 5 - 氧乙酸甲酯，尿嘧啶 - 5 - 氧化乙酸，oxybutoxosine，假尿嘧啶，queosine，2 - 硫胞嘧啶，5 - 甲基 - 2 - 硫尿嘧啶，2 - 硫尿嘧啶，4 - 硫尿嘧啶，5 - 甲基尿嘧啶，和 2，6 - 二氨基嘌呤。

核酸“控制序列”是指转译起始和终止密码子，启动子序列，核糖体结合位点，聚腺苷酸化信号，转录终止序列，上游调节区，加强子，等等，在限定的宿主细胞中对于给定编码序列的转录和转译这些是必要的和足够的。不是所有这些控制序列必须要存在于重组载体中，只要存在那些对所需基因的转录和转译来说是必要的和足够的元件即可。

“可操作地或可操作地连接”是指编码和控制序列的构型可以执行所需的功能。因此，可操作地连接到一个编码序列上的控制序列能够实现该编码序列的表达。在细胞中将一个编码序列可操作地连接到转录调节区或

处于转录调节区控制之下，当 DNA 聚合酶结合到启动子序列上，可将该编码序列转录为 mRNA，该 mRNA 进一步转译为所编码的蛋白质。该编码序列不必定要与编码序列邻接，只要他们具有指导该编码序列的表达即可。因此，例如在启动子序列和编码序列之间可以存在未转译的但转录的间隔序列，并且该启动子序列被认为“可操作地连接”到编码序列上。

“选择性标记”是指任何遗传标记，利用该遗传标记可选择在其基因组中携带该标记的细胞群体。选择性标记的例子包括：营养缺陷型标记，利用这种标记根据在有或没有一种营养物或补充物例如胸腺嘧啶，二氨基庚二酸或生物素的基本培养基上生长的能力选择细胞；代谢标记，利用这种标记根据在含有合适的糖作为唯一碳源的基本培养基上生长的能力或者根据细胞在含有合适的染料或生色底物的培养基上形成带色菌落的能力而选择细胞；以及药物抗性标记，利用这种标记根据在含有一种或多种合适药物，例如，四环素，氨基青霉素，卡那霉素，链霉素或萘啶酮酸的培养基上生长的能力选择细胞。

通常核酸构建体是以转录盒的形式提供的。在构建体中非强制性地包括一个内含子，优选的是 3100bp 并且置于编码序列 5' 末端。通常，优选的是该构建体不会整合到宿主细胞基因组并且该构建体导入到细胞中作为非整合表达盒的组成部分。

用于本发明的构建体包括载体，转录盒，转译盒和质粒。转录和转译起始序列优选地包括转录起始调节区(有时也指“启动子”)和含有“核糖体结合位点”的 5' 未转译序列的转译起始调节区，以便使核糖体结合到 mRNA 并且在起始密码子处驱动转译。优选的是起始控制区的转录和转译功能元件来源于或获自于相同的基因。在一些实施例中通过叠加或缺失序列，或用另一种序列替代可对启动子序列进行修饰。“可获得的”是指与天然基因的那些有足够相似性的调节序列以便为所需 DNA 序列的转录和转译提供所需的特异性。它包括天然和合成序列以及合成和天然序列组合的序列。对于转录起始区，或者启动子元件，可使用任何区域，条件是它提供所需水平的 DNA 序列的转录。转录起始区可以是宿主细胞本身产生的或者是与宿主细胞同源的，和/或是待转录的 DNA 序列产生的或同源的，或者与宿主细胞不相干的或异源的，和/或与待转录的 DNA 序列不相干的或异

源的。“与宿主细胞不相干”是指在宿主中不存在该转录起始区，而含有该转录起始区的构建体将被插入到该宿主中。“与待转录的 DNA 序列不相干的”是指正常情况下该转录起始区不与所需 DNA 序列结合。

位于转录起始调节区下游并且受其控制的是用于插入所需核酸序列的多个克隆位点，所需核酸序列将导致宿主基因型一种或多种变化以及调节宿主表现型。为了方便，可将多个克隆位点以有效的方式用于各种各样的核酸序列。插入到克隆位点的核酸序列具有编码所需多肽的开放式阅读框架，条件是当该编码序列编码所需多肽时，它应该缺失 *crupitic* 的拼接位点，所述位点可能阻碍合适的 mRNA 分子产生和/或产生畸变拼接的或异常 mRNA 分子。该核酸序列可以是 cDNA；也可以是与基因组序列互补的序列，其中基因组序列可以是一个或多个

开放式阅读框架，一个内含子，一个非编码引导序列，或任何其他序列，其中互补序列抑制转录，信使 RNA 的加工，例如拼接或转译。

由于相对而言终止区似乎是可交换的，因此主要使用的终止区是较便利的一种。终止区可能是所需核酸序列本身产生的，或可能是其他来源的。

构建转录载体时，将特定的编码序列定位于具有合适调节序列的载体中，编码序列相对于控制序列的位置和方向应该是编码序列的转录受控制序列的“控制”。为达到该目的对编码特定蛋白质的序列进行修饰是人们所期望的。例如，在某些情况下修饰该序列以便使之以合适的方向结合到控制序列上；或者使之保持阅读框架是必要的。可在插入到载体中之前将控制序列和其他调节序列连接到编码序列上。另一种可选择的方法，将编码序列直接克隆到表达载体中，所述载体已经含有控制序列和位于阅读框架内的合适的限制性位点，所述阅读框架受该控制序列的调节控制。

术语“生化性质等同变异体”是指在某些方面不同于本发明公开的特定的序列的蛋白质或核酸序列，但是，显示具有相同或基本相同的功能。例如，对于 cDNA 的情况，是指含有除特异性公开以外的其他核酸序列的修饰序列，条件是该候补的 cDNA 编码依次编码本发明蛋白质的 mRNA。这样的修饰包括仅几种核酸或更多的核酸的替代。这种修饰涉及简并的编码序列的替代，或用另一种编码序列替代其中的一种编码序列；也包括导入非天然核酸。假如满足功能标准，该候补的 DNA 与本发明公开的核酸杂

交不是必要的。类似地，对于本发明的蛋白质而言，可以使之氨基酸序列发生不影响功能的变化。所述“生化特性等同的突变蛋白”涉及用另一种氨基酸替代一种氨基酸，涉及侧链修饰的或非天然氨基酸的使用和涉及截断。本领域内熟练的技术人员将认识到那一个位点最经得起变化而不影响基本功能。大多数这样的候补序列显示与本发明公开的序列至少有 50 % 的同源性，优选的是，至少 75 %，最优选的是至少 90 %。

用于对 24kb 富伦霉素簇进行测序的方案是从大规模基因组测序使用的方法得到的。将该基因簇片段随机地消化为 400 - 1800bp 片段，从这些片段构建文库用于自动测序。利用 DNA Star 软件在 Macintosh Quandra 上收集序列数据并且操作。

如在限制性图谱(附图 1)中看到的，利用 NotI 可便利地将 24kb 基因簇消化为三个大片段(7.5, 9.0 和 9.5kb)和一个较小片段(1.5 kb)。较小的 NotI 片段位居于富伦霉素生物合成基因簇的 PKS 区的中央并且较大的 NotI 片段位居于两侧。分别将附图 1 中从左到右的各个 NotI 片段亚克隆到 pBluescript 中，得到的克隆分别命名为 Not6, Not7, Not2 和 Not3。将三个较大的 NotI 片段用于产生各自的 MspI 片段文库，实施 shotgun 测序方案。对 NotI 片段的 3' 和 5' 末端测序分析得到的“锚”序列有利于将序列资料与制成酶切图谱的基因簇进行调准。通过在用 MspI 部分消化后进行凝胶电泳分离出 0.4 到 1.8 kb 的片段而从各个 NotI 亚克隆构建随机文库。将这些片段连接到 pBluescript 的 AccI 位点并且转化到合适的大肠杆菌宿主(DH5 α 或 XL 1 - Blue)以便形成单个的文库。将这些文库平板培养并且选择足够量的亚克隆用作为双链测序的模板。

利用 Applied Biosystems(Foster City, CA) Taq dye - 双脱氧终止循环测序反应对这些模板进行测序，该反应利用了聚合酶链反应和荧光标记的核苷酸。读出测序的凝胶，收集数据，用 ABI 373A 自动测序仪分析序列。在获得序列数据后，将他们与以前测定的序列进行校直，以形成重叠的，邻接区域的序列(重叠群)。根据积累的序列数据，将重叠群结合为较大的重叠群。在两条链的大多数区域，进行了三至四次的测序，尽可能提供整个基因簇的确定的序列。

将基因以开放式阅读框架(ORFs)的形式定位于该序列中,该基因中第三位的密码子 G 或 C 出现的频率高,这是链霉菌属基因的一般特性(Bibb MJ,et al.,Gene 30:157(1984))。将以这种方式定位的基因的氨基酸序列与 SWISS - PROT 数据库中的序列进行比较。当 SWISS - PROT 数据库似乎没有某些的序列时,可将推测的该基因的核苷酸序列与 GenBank 数据库中的链霉菌属 DNA 序列进行比较。以这种方式找到其序列与富伦霉素基因簇中基因类似的基因并且指定推测的功能。

富伦霉素基因簇的整个序列描述于 SEQ ID NO: 22 中。下面从该基因簇的左末端开始讨论各个基因。假设的富伦霉素生物合成途径显示于附图 2 中,在该途径中指明了推测的各个基因产物的作用。所注明的基因的功能概括于表 1 中。该基因簇中基因的顺序似乎与生物合成步骤相平行。以前已经报道在大环内酯途径中存在该特征(Donadio S,et al., 1991,Science 252: 675),但在芳香簇聚酮化合物途径中不存在。

据认为基因 A(编码 SEQ ID NO: 1 蛋白质的 SEQ ID NO: 22 的 636 碱基 - 2948 碱基)是该基因簇的第一个基因,因为该基因上游有 1kb 的非编码区并且该基因上游 5' 末端的部分基因与序列数据库中任何序列不同源。基因 A - D 和 F 含有一个亚簇,它可能参与富伦霉素的溢出。基因 A - D 的产物代表 ABC 运输系统的成分。ABC 运输系统特征性特点包括各具有两个保守序列基元的两个 ATP 结合区,和各具有六个跨膜序列的两个膜区(Higgins CF, Ann Rev Cell Biol 8: 67(1992))。两个 ATP 结合区和两个膜区可能都在相同多肽上,如具有真核的多个药物抗性蛋白,在这种情况下多肽由代表内部复制品的两个半个组成(Chen C et al.,1986, Cell 47: 381)。细菌系统经常在不同的多肽上有不同的区,但是几乎在所有情况下膜和 ATP 结合成分被装配成一个多聚复合物,以便在所有运输系统中其三维空间定位是类似的。细菌摄取系统通常在膜的外面具有一个附加成分,该成分结合待运输的底物;溢出系统没有该成分(Reizer J,et al.,Prot Sci 1:1326(1992))。

基因 A 和 B(编码 SEQ ID NO: 2 蛋白质的 SEQ ID NO: 22 的 2945 碱基 - 3916 碱基)和基因 C(编码 SEQ ID NO: 3 蛋白质的 SEQ ID NO: 22 的 4020 碱基 - 4844 碱基)和基因 D(编码 SEQ ID NO: 4 蛋白质的 SEQ

ID NO: 22 的 4841 碱基 - 6415 碱基)的转译结合在一起进行的(指一个基因的终止密码子邻接另一个基因的起始密码子), 这通常是共表达的基因具有精确的化学计量学。基因 D 编码一个可溶性的 524 氨基酸蛋白质, 它含有各具有两个保守的序列基元的两个 ATP 结合区。该基因 D 蛋白质由两个同源的半个组成, 各与推测的波赛链霉菌, DrrB 溢出系统的可溶性成分 (Guilfoyle PG & Hutchinson CR, *Proc Natl Acad Sci* 88: 8553(1991))具有同源性, 有人提出该蛋白质是将 anthracyclines, 阿霉素和道诺红霉素泵出细胞外。该基因 D 产物与来自许多生物体的多种药物抗性泵具有同源性。基因 C 编码一个 274 氨基酸蛋白质, 它具有此类蛋白质的典型的六个推测的跨膜区并且与几种细菌肽透性酶的两种膜成分具有很好的相似性。由基因 C 编码的蛋白质也含有在独立于摄取系统的结合蛋白的膜成分中经常看到的一个基元。与基因 A 结合在一起转译的基因 B 也编码有六个推测的跨膜序列和具有与上面提到相同的细菌肽透性酶的其他膜成分相似的序列。由于基因 A 和 B 结合在一起转译的, 因此可能基因 A 具有某些运输的作用; 但是, 基因 A 与序列数据库中的其他序列没有显著的同源性。

基因 F(编码 SEQ ID NO: 6 蛋白质的 SEQ ID NO: 22 的 7344 碱基 - 8897 碱基)与放线菌紫素 actVA-ORF1 基因(Caballero JL, et al., *Mol Gen Genet* 230; 401(1991))同源并且与 actII-ORF2 基因(Fernandez-Moreno MA, et al., *Cell* 66: 769(1991))几乎没有同源性。有人推测这两类基因参与放线菌紫素的溢出。基因 F 也与 *Nocardia lacamdurans* 的头霉素 C 基因簇的 cmcT 基因(Coque JJR, et al., *EMBO J* 12: 631(1993))类似。因此基因 F 可能与基因 A, B, C 和 D 一起起作用, 或代表了一个独立的不是直接由 ATP 水解驱动的溢出系统。在基因 D 和 F 之间是基因 E(编码 SEQ ID NO: 5 蛋白质的 SEQ ID NO: 22 的 6533 碱基 - 7183 碱基), 它编码一个与数据库中的其他基因不具有显著相似性的 216 氨基酸可溶性蛋白质。

基因 G(编码 SEQ ID NO: 7 蛋白质的 SEQ ID NO: 22 的 9164 碱基 - 10012 碱基)编码与其他聚酮化合物基因簇转录激活子, 包括放线菌紫素基因簇中的 actII-ORF4 基因(Fernandez-Moreno MA, et al., *Cell* 66: 769(1991))和 道诺红霉素基因簇中的 dnrI(Stutzman-Engwall KJ, et

al., *J Bact* 174:144(1992))具有显著相似性的蛋白质。可能它通过与下文中讨论的推测的启动子基元发生相互作用激活该基因簇中的其他许多基因。

基因 H, I, J 和 K 一套基因都位居于一个可能参与二酮化合物起始单位(例如, 丁酰 - ACP, 丁烯酰 - ACP, 羟基丁酰 - ACP 或乙酰乙酰 - ACP)的合成并且将其转移到聚酮化合物合成酶以便启动富伦霉素的生物合成。基因 I(SEQ ID NO: 9)的蛋白质与大肠杆菌的 *fabH* 蛋白质具有显著的同源性, 它催化乙酰 CoA 与丙二酰 - CoA 缩合形成乙酰乙酰 - ACP, 据信, 在 II 型系统中它启动脂肪酸合成(Tsay J-T, et al., 1994, *J Biol Chem* 267: 6807)。尽管在柔毛霉素的基因簇(Ye J, et al., 1994, *J Bacteriol* 176: 6270)和阿霉素基因簇中(Grimm A, et al., 1994, *Gene* 151: 1)中已经看到与 *fabH* 具微弱相似性的基因, 但是这是第一次在芳香簇聚酮化合物基因簇(该基因簇有保守序列基元环绕活性位点半胱氨酸)中发现 *fabH* 同系物。在放线菌紫素基因簇中不存在 *fabH* 同系物。尽管当与其他聚酮化合物合成酶基因在异源系统中共表达时基因 M 似乎提供松散的链长度特异性(McDaniels R, et al., 1993, *J Am Chem Soc* 115: 11671), 但是本发明讨论的玫瑰暗黄色链霉菌菌株仅产生 18 碳富伦霉素而不是 16 碳结构例如 七尾霉素或 卡拉真菌素。因此, 基因 I(编码 SEQ ID NO: 9 蛋白质的 SEQ ID NO: 22 的 11628 碱基 - 10618 碱基)可能在启动富伦霉素的生物合成和确保 18 碳长度中具有关键的作用。与基因 I 结合在一起转译的是基因 H(编码 SEQ ID NO: 8 蛋白质的 SEQ ID NO: 22 的 10621 碱基 - 10105 碱基), 它与各种氧化还原酶基因和与放线菌紫素基因簇的 *actI*-ORF2 区域具有微弱的相似性。基因 H 的产物可能参与乙酰乙酰 - ACP 还原为丁酰 - ACP。

基因 J(编码 SEQ ID NO: 10 蛋白质的 SEQ ID NO: 22 的 11809 碱基 - 12066 碱基)编码不同于基因 N 产物(见下文)的酰基载体蛋白, 基因 I 的蛋白质(与基因 H 蛋白质一起)利用该产物形成起始单位。基因 K(编码 SEQ ID NO: 11 蛋白质的 SEQ ID NO: 22 的 13209 碱基 - 12154 碱基)蛋白质与来自阿霉素(Grimm A, et al., 1994, *Gene* 151: 1)和柔毛霉素 (Ye J, et al., 1994, *J Bacteriol* 176: 6270)的基因簇的两个推测的酰基转移酶同源。他们都含有酰基转移酶典型的活性位点基元(GlyHisSer)。放线

菌紫素基因簇不含有基因 I 或基因 K 的同源区。

基因 L, M 和 N 编码核心的聚酮化合物合成酶复合物并且与放线菌紫素基因簇的 actI 基因(Fernandez-Moreno MA, et al., 1992, J Biol Chem 267: 19278)同源。基因 L(编码 SEQ ID NO: 12 蛋白质的 SEQ ID NO: 22 的 13409 碱基 - 14686 碱基)蛋白质含有 β -酮酰基合成和酰基转移的活性位点基元。基因 M(编码 SEQ ID NO: 13 蛋白质的 SEQ ID NO: 22 的 14767 碱基 - 16047 碱基)与基因 L 同源, 但是没有活性位点基元。它是功能性聚酮化合物合成酶复合物所必须的。基因 N(SEQ ID NO: 22 的 16120 碱基 - 16371 碱基)编码 SEQ ID NO: 14 蛋白质, 它是一个酰基载体蛋白。基于这些基因的近似性, 似乎聚酮化合物合成酶利用了基因 N 酰基载体蛋白, 其中 fabH 同系物利用基因 J 酰基载体蛋白。

基因 O, P, Q, 和 R 组成了第四个亚簇。基因 O(编码 SEQ ID NO: 15 蛋白质的 SEQ ID NO: 22 的 16935 碱基 - 16453 碱基)与放线菌紫素基因簇的 actVI-ORFA 基因(Fernandez-Moreno MA, et al., 1992, J Biol Chem 269: 24854)同源, 有人认为它催化涉及半缩酮结构的形成和脱水。基因 P(SEQ ID NO: 22 的 17088 碱基 - 17903 碱基)编码与放线菌紫素基因簇的 actIII 酮基还原酶(Hallam SE, et al., 1988, Gene 74:305)同源的 SEQ ID NO: 16 的蛋白质。基因 Q(SEQ ID NO: 22 的 17903 碱基 - 18898 碱基)编码与放线菌紫素基因簇的 actVII 环化酶/脱水酶(Fernandez-Moreno MA, et al., 1992, J Biol Chem 267: 19278)同源的 SEQ ID NO: 17 的蛋白质。有人认为它催化第一个环化反应。基因 R(SEQ ID NO: 22 的 18895 碱基 - 19839 碱基)编码与放线菌紫素基因簇的 actIV 蛋白质, 一种参与第二个环化反应的环化酶/脱水酶(Fernandez-Moreno MA, et al., 1992, J Biol Chem 267: 19278)高度相似的 SEQ ID NO: 18 的蛋白质。由于基因 R 与基因 Q 结合一起转译, 而基因 Q 与基因 P 结合一起转译, 因此这三个基因在一个转译单位内。

基因 S(编码 SEQ ID NO: 19 蛋白质的 SEQ ID NO: 22 的 20907 碱基 - 19990 碱基)显示与一定区域内的 actVA-ORF4, actVI-ORF1 和 actVI-ORF2 基因具有微弱的相似性。有人认为 actVI-ORF1 和 actVI-ORF2 基因产物可分别催化立体特异性酮还原和立体特异性烯酰还原(Fernandez-

Moreno MA, et al., 1992, J Biol Chem 269: 24854)。由于基因 T(编码 SEQ ID NO: 20 蛋白质的 SEQ ID NO: 22 的 22094 碱基 - 20904 碱基)高度类似于 actVA-ORF5(Caballero JL, et al., 1991, Mol Gen Genet 230: 401), actVA-ORF5 催化 actVA 产物的水解(Cole SP, et al., 1987, J Antibiot 40:340), 基因 T 可能编码富伦霉素生物合成中的水解酶类似物。因此, 基因 S 可能编码上文提到的一个或两个立体特异性还原。

在放线菌紫素和富伦霉素基因簇中推测的启动子基元(由 7 个碱基精确地分隔开的两个或多个 TGCA 序列)位于几个推测的操纵子的第一个 ORF 的起始密码子上游 20 - 50bp。该基元仅存在于 isochromanone 生物合成基因簇的推测的启动子区域表明它可能代表参与转录激活的因子的结合位点。在含有该基元的放线菌紫素基因簇中从启动子开始转录的基因已知是由 actII-ORF4 基因产物调节的(Gramajo HC, et al., 1993, Mol Microbiol 7:837)。因此, 富伦霉素基因簇中的基因 G(它与放线菌紫素基因簇中的 actII-ORF4 基因同源)很可能是参与激活含有该基元的富伦霉素基因簇中的启动子的调节因子。含有该基元的许多启动子可能提供双向转录。

在该基因簇的右末端的基因与来自各种各样的生物体编码甘油醛 - 3 - 磷酸脱氢酶(G3PDH)的基因高度类似。该基因簇中不可能定位 G3PDH 基因, 因此它可能定位于该基因簇的右边界。但是, 可观察到(Frohlich K-U, et al., 1989, J Bacteriol 171:6696)在沙场链霉菌生产 pentalenolactone 过程中, 对 pentalenolactone 敏感的 G3PDH 被对该产物有抗性的遗传上不同的异构体替代。因此, 可能与 pentalenolactone 类似, 富伦霉素也抑制 GAPDH, 并且在生产过程中也产生抗性异构体。

在经过测序的富伦霉素基因簇的区域没有基因明显地显示编码脱氧富伦霉素环氧酶或富伦霉素 B 还原酶。可能编码这些活性物质之一的基因仅是 A 和 E, 他们不与数据库中的任何序列匹配并且其功能未知。富伦霉素对链霉菌属和其他细菌有毒, 富伦霉素 B 可能是最毒的, 脱氧富伦霉素低毒, 环氧化富伦霉素毒性最低。因此, 如果从富伦霉素基因簇中失去富伦霉素 B 还原酶或脱氧富伦霉素环氧酶活性, 由此产生的产物将毒害转化子。

为了在玫瑰暗黄色链霉菌和其他链霉菌中使用富伦霉素基因簇基因,构建几个大肠杆菌/链霉菌属穿梭载体。为了在链霉菌属中进行复制,这些载体使用了来自质粒 pSG5(Muth G et al., 1988,Mol Gen Genet 211: 424; Muth G et al., 1989,Mol Gen Genet 219: 341)的温度敏感性复制子,该复制子在 *S.ghanaensis* 的基因组中其拷贝数约为 15 并且具有较宽的宿主范围。这些复制子的明显的优点是在升高的温度(35 ℃ 而不是 28 ℃)下转化子生长时阻止了自主复制,从而可以选择质粒与染色体的整合,这是提供用于工业生产的遗传稳定的菌株必须的。这些载体在大肠杆菌中繁殖的主链是石蕊 28 或 38(New England Biolabs,Beverly,Mass.,USA),它是以 pUC 载体为基础的。由于已经显示利用单链 DNA 转化链霉菌宿主比用双链形式的相同载体可以将特异性整合的频率提高 10 - 100 倍(Hillemann D,et al.,1991,Nucleic Acids Res 19:777),可以使用石蕊载体,因为他们包括一个 M13 源点,可以通过用 M13 辅助噬菌体感染大肠杆菌而使单链 DNA 分离。

为了使本发明的 DNA 序列尤其是来自富伦霉素基因簇的基因(或转录单位)过量表达,将 DNA 序列,基因或操纵子连接到链霉菌强启动子特别是 *tipA* 和 *ermE*^{*} 启动子(附图 4)上。为此,利用提供所需限制性位点的引物进行 PCR 扩增所需的基因或区域,将 PCR 产物克隆到合适的载体 pErme 或 pTipA(附图 5)上。然后利用限制性位点方便地将各个启动子-基因嵌合构建体移动到上文描述的穿梭载体,所述限制性位点重新构建了两个半个的选择性标记(例如,利用氨苄青霉素抗性基因中的独特的 *AseI* 位点)。所需的基因或区域也可由其本身的启动子表达,通过修饰启动子或通过导入多拷贝的构建体可以获得较高水平的表达。如果这些基因的过量表达导致更高的富伦霉素效价,则通过在较高温度下保持选择压力消除质粒,通过强制发生同源性整合这些载体可产生稳定生产的菌株。

在链霉菌中用于起动转录的有效启动子元件包括 *tipA* 和 *ermE*^{*},以及来源于已知以高水平表达的许多基因的启动子。利用各种各样的标准技术包括在预定的位置制造独特的选择性位点的位点特异性诱变技术(Kunkel TA,1985, *Proc. Natl. Acad. Sci. U.S.A.*, 82:488-492),合成寡核苷酸接头,或利用制备所需限制性位点的引物进行 PCR 扩增,可将调节序列连

接到所需基因。在理想的实施例中，将所需序列连接到该调节序列上以便在基因的#1 密码子起动转译。

用于重组生产富伦霉素的宿主细胞可来源于具有携带本发明重组富伦霉素基因簇或相应 DNA 序列能力的任何生物体。因此，本发明的宿主细胞可来源于原核或真核生物体。但是，优选的宿主细胞是那些从放线菌构建得到的，放线菌是一类许多聚酮化合物的丰富生产者的丝状细菌。用于本发明的特定的优选的属是链霉菌属。例如，生二素链霉菌，生金色链霉菌，*S. avermitilis*，天蓝色链霉菌 *S. azureus*，肉桂地链霉菌，天蓝色链霉菌，古希腊链霉菌，红霉素链霉菌，弗氏链霉菌，加利利链霉菌，淡青链霉菌，吸水链霉菌，淡紫青链霉菌，小小链霉菌，波赛链霉菌，龟裂链霉菌，麦玫瑰暗黄色链霉菌，嗜热链霉菌，紫红链霉菌，及其他，可提供用于本发明的便利的宿主细胞，而玫瑰暗黄色链霉菌是优选的。参见，例如，Hopwood, D. A. 和 Sherman, D. H. *Ann Rev Genet* (1990) 24:37-66；O'Hagan, D. *The Polyketide Metabolite* (Ellis Horwood Limited, 1991)，描述了各种产生聚酮化合物的生物体和其天然产物。例如从任何已知的保藏库，例如美国典型培养物保藏中心 (ATCC)，Centraalbureau voor Schimmelcultures (CBS) 或德国微生物和细胞培养物收集处 GmbH (DSM) 或在杂志 “Industrial Property” [(1991) 1, pp29-40] 列出的其他保藏机构获得的所述宿主也可用于分离富伦霉素基因簇 DNA 序列。例如玫瑰暗黄色链霉菌可从 ATCC 获得，保藏号为 ATCC19921 或 ATCC19805 或从 CBS 获得，保藏号为 CBS577.68。在该文中应该记住基于本文描述的序列信息，利用本领域内已知的方法或描述于例如 EP747 483 的方法，也可合成制备本发明的 DNA 序列。

利用标准的方法例如通过同源重组，缺失天然存在的 PRS 基因可用遗传工程方法制备上文描述的细胞(参见，例如，Khosla, C. et al., 1992, *Molec. Microbiol.* 6:3237)。利用本领域技术人员已知的方法可将集合在一起编码替代的富伦霉素基因簇的基因序列或本发明的 DNA 序列或其功能等同物插入到一种或多种表达载体中。表达载体包括可操作连接到所需的富伦霉素编码序列上的控制序列。用于本发明的合适的表达系统包括在真核和原核宿主细胞起作用的系统。但是，如上文解释的，优选的是原核系

统，特别是与链霉菌相容的系统是特别有利的。用于所述系统的控制元件包括启动子，非强制性地含有操纵子序列，和核糖体结合位点。特别优选的启动子包括来源于 PKS 基因簇的控制序列，例如在放线菌紫素和富伦霉素基因簇的转译单位的 5' 存在的那些。但是，其他启动子，例如来源于以高水平表达的基因的启动子也可用于本发明的构建体中。例子包括来源于编码分解代谢酶(蛋白酶或糖苷酶)，生物合成酶(例如在色氨酸操纵子中的那些)，抗菌素抗性蛋白(例如 β - 内酰胺酶)或噬菌体结构蛋白的基因的启动子序列。另外，在自然界中不存在的合成启动子，例如 tac 启动子(美国专利 No.4, 551, 433)也可用于本发明的构建体中。其他调节序列也是令人满意的，可用于调节富伦霉素替代序列相对宿主细胞的生长而表达。调节序列是本领域内技术人员已知的，并且举例包括引起基因的表达对化学或物理刺激物包括存在的调节化合物反应而开启或关闭的那些。其他类型的调节元件也可存在于载体例如，那些来自于在特定生长期间表达的基因的元件。在重组表达载体中包括各种可选择的标记。已知许多标记物可用于选择转化细胞，并且通常含有一个基因，当细胞在合适的选择性培养基上生长时该基因表达后给转化细胞提供可选择性表现型。所述标记包括例如提供抗菌素抗性的基因或能够在基本培养基上生长。

可将所需的各种富伦霉素生物合成酶基因(和其功能性亚单位)或本发明的 DNA 序列克隆到一个或多个重组载体上，以便他们分别处于各自的启动子控制之下。另一种可选择方法，将基因组成为处于例如单个启动子控制之下的操纵子(他们天然存在于富伦霉素基因簇中或合成制备的)。功能性亚单位序列可包括侧面的选择性位点以便其他亚单位容易缺失和/或插入，从而可以制备杂合的 PKSs。所述独特的限制性位点的设计对本领域内技术人员来说是已知的并且可利用上文描述的技术例如定位诱变技术和 PCR 来完成。用于将本发明的重组载体导入到合适的宿主中的方法对本领域内技术人员来说是已知的。对于链霉菌，通常通过用溶菌酶消化细胞壁制备原生质体，并且用转化 DNA 与各种刺激其吸收的试剂例如聚乙二醇和二价阳离子来处理原生质体(Hopwood DA, et al., 1985, Genetic Manipulation of Streptomyces-A Laboratory Manual, John Innes Foundation, Norwich, UK)。随后使原生质体再生为完整细胞并且选择转化子。也可以采用电穿孔

的方法将 DNA 导入到链霉菌和其他细菌中。一旦 PKS 基因被表达, 可以利用已知技术鉴别产生聚酮化合物的菌落并且分离。由转化子产生的聚酮化合物可在合适的条件下进一步被加工。

由 Iwai et al., J. Antibiotics 31: 959(1987)和 Omura et al. 在美国专利 No.4, 199, 514 上描述了发酵的典型方案。可在合适的碳源和氮源例如, 鱼肝油, 玉米油, 葡萄糖, 麦芽糖等等作为碳源, 黄豆粉, 蛋白胨, 酵母等等作为氮源中培养用本发明载体转化的微生物例如玫瑰暗黄色链霉菌。通过好氧条件下, 20 - 35 °C 培养约 100 - 150 小时完成发酵。将培养肉汤与细胞分离并且过滤, 用亲水有机溶剂例如乙酸乙酯, 丁酸乙酯, 苯等等从滤液提取富伦霉素。然后采用常规技术例如凝胶层析将富伦霉素互相分离并且中间产物化合物(环氧和脱氧富伦霉素)转化为富伦霉素 B。另一种可选择的方法, 在大气压下, 室温, pH8 - 10 条件下, 将无细胞发酵液氢化以便将环氧化物转化为脱氧富伦霉素。在氢化之后, 将 pH 调节到 5.5 并且在 60 °C 向反应混合物中吹入空气以便将脱氧富伦霉素 转化为富伦霉素 B。然后提取富伦霉素 B 并且沉淀。

实施例 1

附图 3 显示链霉菌/大肠杆菌穿梭载体 pSSVtsr 和 pSSVaph 的构建。从载体 pGM160(Muth et al., 1989)分离表达盒形式的 pSG5 链霉菌复制子和硫链丝菌肽抗性基因,通过缺失 HindIII 插入物形式的转座子可从载体 pCZA168(Solenberg 和 Baltz, 1991)获得 pGM160 本身。将 pGM160 用 BamHI 完全消化和用 BclI 部分消化, 凝胶分离出含有 pSG5 链霉菌复制子的 2.6kb 片段并且连接到石蕊 28 和石蕊 38 (New England Biolabs)。重新构建得到最接近于在石蕊载体的多接头中的 SnaBI 的 BamHI 位点的 SG5 插入物的定位方向的选择应该得到质粒 pL28SG 5 和 pL38SG5。按照 Muth et al., 1988 描述, 基于对基本复制子 pSG5 的测定, 按如下所述, 将非必要 DNA 序列缺失以便得到 pL28SG 5a 和 pL38SG5a。

用 KsI 消化 pL38SG5 并且将凝胶分离得到的 4.8kb 片段重新连接得到 pL38SG5a。用 KsI 和 BamHI 消化 pL38SG5 并且将凝胶分离得到的 4.8kb 片段连接到下文显示的合成接头得到 pL38SG5a。

GATCCGCAGTACTGCC

GCGTATGACGGCGCG

从 pCZA168 分离出 1.1kb BclI/EcoRI 片段形式的硫链丝菌肽抗性标记并且将其克隆到石蕊 28 的 BamHI/EcoRI 位点得到 pL28tsr。从 pIJ680(Hopwood, et al., 1985)分离 1.0kb SaccI 片段形式的弗氏链霉菌的新霉素抗性基因(aph)并且将其克隆到 pBluescript SK-(Stratagene, La Jolla, CA, USA)得到 pBSaph。

用 NdeI 切割 pL28tsr 并通过用大肠杆菌 DNA 聚合酶 I 的 Klenow 片段填充位点制备平齐末端。然后将该载体用 HindIII 切割, 凝胶分离得到的 1.1kb 片段。用 HindIII 和 HpaI 切割 pL38SG5a, 凝胶分离 5.2 kb 片段。然后将这两个片段连接得到 pSSVtsr。通过用 BamHI 和 Bgl II 切割并且重新连接从 pSSVtsr*得到 pSSVtsr。

用 SpeI 切割 pBSaph, 通过用大肠杆菌 DNA 聚合酶 I 的 Klenow 片段填充使位点成平齐末端。然后用 Tsp509I 切割该载体, 凝胶分离出 1.1kb 片段。用 EcoRI 和 HpaI 切割 pL38SG5a, 凝胶分离 5.2kb 片段。然后将这两个片段连接得到 pSSVaph。

实施例 2

用方便的限制性位点合成 tipA 和 ermE* 启动子。在 ABI 392 DNA 合成仪上合成一系列重叠的寡核苷酸并且随后用 T4 多核苷酸激酶使之磷酸化。然后通过加热到 90 °C 并且慢慢冷却而使这些寡核苷酸退火。在退火之后, 室温下用 T4 连接酶将这些片段连接 4 小时并且然后用 KpnI 和 BamHI 切割。将凝胶分离得到的片段克隆到石蕊 28 的 KpnI 和 BamHI 位点得到 pTipA 和 pErmE(附图 5)。

通过 PCR 扩增富伦霉素基因簇中的各种基因, 克隆到 pTipA 和/或 pErmE, 然后采用基本上与描述用于基因 G 的转录激活子(下文)相同步骤转移到穿梭载体(附图 5)。按下文的描述构建用于过量表达选定的基因的载体。利用下文显示的引物将基因 G 进行 PCR 扩增。用 BamHI 消化得到的 PCR 产物(对于待连接到 pTipA 的产物的情况要加 NdeI), 凝胶分离, 并且克隆到 pErmE 的 BamHI 位点或 pTipA 的 BamHI 和 NdeI 位点之间分别得到 pErmTA 和 pTipTA。

向前(5') PCR 引物, 用于扩增与 ermE* 融合的基因 G

CGGGATCCAGCGGTGGAGATCAAGTACATGGGTCAGTTGACC

向前(5') PCR 引物, 用于扩增与 tipA 融合的基因 G

CGCATATGGAGATCAAGTACATGGGTCAGTTGACC

逆转(3') PCR 引物, 用于扩增克隆到 pErmE 或 pTipA 的基因 G

GCGGATCCGTGTCAGTCGTGCGAGCGCGCCGCGGTGGC

用 AseI 和 StuI 切割 pErmTA 和 pTipTA 载体并且凝胶分离较大(2.8kb)片段。在同时, 用 AseI 和 SnaBI 切割 pSSVtsr 和 pSSVaph 载体, 凝胶分离较大的(3.9kb)片段。将 pErmTA 和 pSSVtsr 片段连接在一起得到 SSVtsr-ermTA。将 pTipA 和 pSSVaph 片段连接在一起得到 SSVaph-tipTA。为了借助于 TipA 启动子构建表达基因, 利用新霉素抗性基因选择, 以便利用硫链丝菌肽进行诱导。由于富伦霉素基因簇中缺失 AseI, SnaBI, StuI, BamHI 和 NdeI 位点, 利用该方案可表达任何富伦霉素基因。

为了制备设计为允许该基因簇中的特定基因分离的载体, 将该基因内部的片段克隆到石蕊 28 或 38, 并且利用同样的方案将片段转移到穿梭载体中。例如, 将跨越大部分基因 L 和基因 G 5' 末端的 1.6kb 的 Sal I 片段连接到石蕊 38 并且将得到的载体与 pSSVtsr 连接得到 SSVtsr-deltaLM。

实施例 3

将用于转化链霉菌的载体在甲基化缺陷型(dam-)(dcm-)大肠杆菌菌株中繁殖。从该菌株分离到的载体 DNA 用于转化变青链霉菌, 天蓝色链霉菌, 玫瑰暗黄色链霉菌。用以前描述的修饰步骤(Hopwood DA, et al., 1985, Genetic Manipulation of Streptomyces-A Laboratory Manual, The John Innes Foundation, Norwich, UK)转化包括玫瑰暗黄色链霉菌的链霉菌属。用孢子悬浮液接种到一个 250ml 带隔板的烧瓶中, 该烧瓶中有 25ml 的改良的 YEME 培养基(每升: 3g Difco 酵母抽提物, 5g Difco 细菌用蛋白胨, 3g Difco 麦芽抽提物, 10g 葡萄糖, 250g 蔗糖; 在高压蒸汽灭菌之后将 MgCl₂ 和甘油分别加入至 5mM 和 0.1% 终浓度)并且在 30℃ 以 220RPM 的转速振荡培养 5 - 8 天。得到的菌丝体在 10.3% 蔗糖中洗二次, 重新悬浮于 4ml 裂解缓冲液(Hopwood et al., 1985, ibid.)中, 该缓冲液含有 0.5mg/ml 溶菌酶和 0.5mg/ml 无色肽 (achromopeptides)。在室温下使

细胞壁消化 10 - 15 分钟并且用显微镜监测。通过将消化产物透过棉布塞子而除去留下的菌丝体，将原生质体在新鲜原生质体缓冲液(Hopwood et al.,1985, *ibid.*)中洗涤二次并且储存于 - 80 ℃。

在室温下通过在含有 10 μ l 痕量元素和 500 μ l 溶于转化缓冲液(Hopwood et al.,1985, *ibid.*)的 25 % PEG6000 的溶液中用 0.5-5 μ g 质粒 DNA 保温 50 μ l 冻融的原生质体样品完成转化 1 分钟。将转化物与 500 μ l 原生质体缓冲液混合并且将 100 - 500 μ l 样品平铺在再生平板(CM1-2,含有 10.3%蔗糖)上。在 30 ℃培养 24 小时之后用含有硫链丝菌肽(至终浓度为 30mg/l)或新霉素(至终浓度为 10mg/l)的软琼脂培养基覆盖在平板上，并且分离得到的转化细胞。

表 1

<u>基因(SEQ ID NO)</u>	<u>所编码的蛋白质的特性</u>
A(SID: 1)	80kD, 没有高度同源性的非膜蛋白
B(SID: 2)	与 A 结合一起转录的膜蛋白; 可能与运输有关
C(SID: 3)	具有六个跨膜区的蛋白质
D(SID: 4)	与 C 结合一起转译的 ABC 运输系统的可溶性 ATP 结合成分
E(SID: 5)	功能未知; 在数据库中无同源序列; 不是膜蛋白
F(SID: 6)	与 actVA-ORF1 相关的推测的溢出泵; 12 - 14 跨膜区
G(SID: 7)	与 actII-ORF4 相关的转录激活子
H(SID: 8)	与基因 I 结合一起转译
I (SID: 9)	fabH 的同系物; 可能合成四碳起始物单位
J(SID: 10)	可能由 fabH 的同系物利用的 ACP 以便起始物单位
K(SID: 11)	酰基转移酶; 可从 ACP(J)上将起始物单位转移到 PKS
L(SID: 12)	PKS 酮酰基合成酶亚单位
M(SID: 13)	其他的 PKS 亚单位; 称之为“链长度因子”
N(SID: 14)	由 PKS 利用的 ACP
O(SID: 15)	与 actVI-ORFA 相关; 可能是一个半缩酮脱氢酶
P (SID: 16)	与 actIII 相关的酮基还原酶
Q(SID: 17)	与 actII 相关的环化酶/脱氢酶
R(SID: 18)	与 actIV 相关的环化酶/脱氢酶
S(SID: 19)	与 actVI 因子微弱相关的氧化还原酶
T(SID: 20)	与 actVA-ORF5 相关的形成苯醌的羟化酶

序 列 表

(1)一般资料

(i)申请人:

(A)姓名: F.HOFFMANN-LA ROCHE AG

(B)街道: Grenzacherstrasse 124

(C)城市: Basle

(D)州: BS

(E)国家: 瑞士

(F)邮政编码(ZIP): CH - 4002

(G)电话: 061 - 688 25 05

(H)传真: 061 - 688 13 95

(I)电传: 962292/965542 hlr ch

(ii)发明名称: 富伦霉素基因簇

(iii)序列数: 22

(iv)计算机可读形式:

(A)介质类型: Floppy disk

(B)计算机: IBM PC 兼容机

(C)可操作系统: PC - DOS/MS-DOS

(D)软件: PatentIn Release #1.0 , 版本#1.30

(2)SEQ ID NO: 1 的资料:

(i)序列特征:

(A)长度: 770 个氨基酸

(B)类型: 氨基酸

(C)链数: 单链

(D)几何结构: 线性

(ii)分子类型: 蛋白质

(iii)假设的: 不是

(iv)反意义: 不是

(ix)序列描述: SEQ ID NO: 1

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

Gly Ala Thr Phe Ser Asp Gly Thr Pro Val Asp Ala Arg Ala Val Lys
 355 360 365
 Ala Asn Phe Asp Ala Val His Glu Leu Gly Ala Ala Ala Ser Arg Gly
 370 375 380
 Ala Val Tyr Leu Asp Gly Tyr Arg Glu Thr Arg Val Ser Gly Ala Arg
 385 390 395 400
 Thr Leu Thr Val Val Phe Asp Lys Pro Asn Ala Gln Phe Leu Arg Gly
 405 410 415
 Thr Ser Thr Val Ser Leu Gly Leu Leu Ser Pro Gly Ser Leu Arg Arg
 420 425 430
 Thr Pro Gln Glu Arg Cys Thr Gly Arg Leu Val Gly Ser Gly Pro Phe
 435 440 445
 Val Leu Asp Arg Tyr Arg Pro Asn Thr Ser Val Thr Leu Asp Arg Arg
 450 455 460
 Lys Gly Tyr Ser Trp Gly Ser Arg Leu Trp Gln Arg Glu Gly Gly Ala
 465 470 475 480
 Tyr Leu Glu Gly Val Glu Tyr Arg Ile Val Pro Glu Asn Thr Thr Arg
 485 490 495
 Ser Gly Ala Leu Ser Ala Gly Gln Leu Asp Val Ala Thr Ala Leu Ala
 500 505 510
 Pro Gln Asp Arg Glu Arg Phe Ser Ala Pro Gly Trp Ser Leu Leu Thr
 515 520 525
 Arg Thr Ala Pro Gly Val Asp Leu Ser Leu Tyr Val Asn Ala Arg Arg
 530 535 540
 Thr Ala Leu Arg Glu Ala Ala Val Arg Gln Ala Leu Gln Lys Gly Ile
 545 550 555 560
 Asp Arg Glu Ala Val Ala Thr Thr Phe Leu Ser Ser Arg Lys Leu Ala
 565 570 575
 Ala Thr Ser Val Leu Ser Ser Thr Thr Pro Gly Tyr Thr Asp Leu Gly
 580 585 590
 Asp Arg Leu Ala His Asp Pro Ala Gly Ala Arg Arg Leu Leu Asp Lys
 595 600 605
 Ala Gly Trp Arg Pro Gly Ala Asp Gly Ile Arg Val Lys Asn Gly Val
 610 615 620
 Arg Leu Arg Leu Asp Ala Val Phe Val Arg Gln Gln Ser Leu Glu Leu
 625 630 635 640
 Val Gln Gln Gln Leu Lys Asp Ile Gly Val Glu Leu Arg Leu Arg Gln
 645 650 655
 Leu Thr Val Ser Arg Phe Pro Glu Val Leu Ala Ala Gly Ser Tyr Asp
 660 665 670
 Leu Ser Leu Gln Ser Ala Asn Arg Ala Asp Pro Asp Val Leu Thr Thr
 675 680 685

Ala	Phe	Ala	Gly	Gly	Thr	Pro	Val	Ala	Asp	Ala	Arg	Leu	Arg	Ser	Glu	
690						695					700					
Leu	Arg	Arg	Ala	Thr	Ser	Ser	Thr	Asp	Glu	Ala	Thr	Arg	Ser	Ser	Leu	
705					710					715					720	
Phe	Ala	Ala	Ala	Gln	Arg	Arg	Leu	Val	Asp	Glu	Gly	His	Val	Leu	Pro	
				725					730					735		
Leu	Asn	Glu	Thr	Glu	Glu	Thr	Ala	Ala	Leu	Ser	Thr	Arg	Val	His	Gly	
			740					745					750			
Leu	Thr	Arg	Asp	Ala	Ser	Asn	Arg	Leu	Val	Leu	His	Asp	Thr	Trp		
		755					760					765				
Thr	Thr	Gly														
		770														

(2) SEQ ID NO: 2 的资料

(i) 序列特征:

(A) 长度: 323 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 2

Met	Thr	Ala	Arg	Tyr	Leu	Ala	Arg	Arg	Leu	Gly	Arg	Val	Val	Leu	Val	
1				5					10					15		
Val	Trp	Ala	Ala	Tyr	Thr	Leu	Ser	Phe	Ala	Val	Leu	Tyr	Leu	Leu	Pro	
		20						25					30			
Gly	Asp	Pro	Val	Gln	Thr	Met	Leu	Ser	Gly	Ala	Ala	Gly	Gly	Asp	Gly	
		35					40					45				
Ala	Ala	Val	Asp	Pro	His	Glu	Ala	Gln	Arg	Leu	Arg	His	Thr	Leu	Gly	
		50				55				60						
Leu	Asp	Arg	Pro	Leu	Ala	Val	Gln	Tyr	Thr	Ser	Met	Leu	Gly	His	Ala	
65				70						75				80		
Leu	Arg	Gly	Asp	Leu	Gly	Thr	Ser	Ile	Arg	Ser	Gly	Ala	Pro	Val	Arg	
			85						90					95		

Gly	Gln	Leu	Ala	Gln	Ala	Leu	Pro	Asp	Thr	Leu	Ser	Val	Ala	Leu	Pro	100	105	110	
Ala	Leu	Val	Leu	Ser	Val	Leu	Val	Ala	Leu	Cys	Leu	Ala	Leu	Leu	Gly	115	120	125	
Ala	Trp	Pro	Arg	Arg	Arg	Ala	Leu	Arg	Arg	Ala	Ala	Thr	Ala	Leu	Pro	130	135	140	
Ser	Leu	Gly	Thr	Ala	Met	Pro	Ser	Phe	Trp	Leu	Gly	Leu	Leu	Leu	Ala	145	150	155	160
Gln	Trp	Val	Ser	Phe	Arg	Trp	Gly	Leu	Leu	Pro	Ala	Thr	Gly	Gly	Gly	165	170	175	
Arg	Ser	Pro	Arg	Ala	Thr	Leu	Leu	Ala	Ala	Leu	Thr	Leu	Ala	Leu	Pro	180	185	190	
Ile	Gly	Cys	Val	Leu	Ala	Gln	Val	Leu	Gly	Arg	Gly	Leu	Arg	Ala	Ala	195	200	205	
Leu	Ala	Glu	Pro	Tyr	Ala	Asp	Val	Ala	Arg	Ser	Arg	Gly	Ala	Gly	Arg	210	215	220	
Ala	Arg	Leu	Leu	Leu	Ala	Arg	Ala	Leu	Arg	Asn	Ala	Ser	Val	Ala	Ala	225	230	235	240
Leu	Ala	Leu	Leu	Gly	Val	Val	Cys	Gly	Gln	Leu	Leu	Ala	Gly	Ala	Val	245	250	255	
Leu	Val	Glu	Thr	Val	Phe	Ala	Arg	Gly	Gly	Ile	Gly	Arg	Leu	Ala	Met	260	265	270	
Asp	Ala	Val	Thr	Tyr	Gln	Asp	Leu	Pro	Val	Val	Gln	Gly	Val	Val	Val	275	280	285	
Leu	Ala	Ala	Leu	Val	Ala	Ala	Leu	Val	Asn	Leu	Val	Val	Asp	Leu	Leu	290	295	300	
Leu	Pro	Leu	Leu	Glu	Pro	Arg	Thr	Ala	Ser	Glu	Ala	Ala	Asp	Ala	Val	305	310	315	320
Pro Ala His																			

(2) SEQ ID NO: 3 的资料

(i) 序列特征:

(A) 长度: 274 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 3

```
Met Ala Leu Arg Arg Val Ala Ala Leu Trp Arg Ala Pro Gly Leu Ala
1          5          10          15
Leu Ser Leu Leu Val Leu Leu Leu Val Leu Gly Trp Ala Leu Leu Pro
20          25          30
Gly Leu Phe Thr Ala Ala Asp Pro Leu Arg Ala Asp Ala Ala His Arg
35          40          45
Leu Leu Ala Pro Gly Ala Gly His Pro Leu Gly Ala Asp His Val Gly
50          55          60
Arg Asp Leu Tyr Ala Arg Val Val His Gly Thr Ala Arg Ser Leu Gly
65          70          75          80
Thr Ala Phe Ala Ala Val Ala Leu Gly Val Leu Ala Gly Gly Ala Leu
85          90          95
Gly Ala Val Ala Gly Val Ala Gly Arg Ala Val Asp Ala Val Val Met
100          105          110
Arg Val Val Asp Val Leu Leu Ala Val Pro Gly Leu Leu Leu Ser Leu
115          120          125
Ala Val Val Ser Ala Leu Gly Phe Gly Thr Ala Gln Val Ala Cys Ala
130          135          140
Val Gly Val Gly Thr Val Gly Gly Ile Ala Arg Val Ser Arg Ala Gln
145          150          155          160
Val Arg Arg Val Arg Gly Gly Glu Tyr Val Glu Ala Ala Arg Leu Ala
165          170          175
Gly Val Ala Gly Pro Leu Ile Leu Leu Arg His Ile Val Pro Asn Ala
180          185          190
Ala Pro Pro Val Leu Ala Leu Ala Val Thr Glu Cys Gly Thr Ala Val
195          200          205
```

Leu	Gly	Val	Ala	Ser	Leu	Gly	Phe	Leu	Gly	Phe	Gly	Ala	Pro	Pro	Pro
210						215					220				
Ala	Pro	Glu	Trp	Gly	Ala	Leu	Ile	Ser	Thr	Gly	Arg	Asp	Tyr	Leu	Val
225					230					235				240	
Ser	Ala	Trp	Trp	Leu	Thr	Thr	Leu	Pro	Gly	Leu	Val	Leu	Val	Ala	Leu
				245					250					255	
Val	Val	Ala	Leu	His	Arg	Val	Gly	Arg	Ala	Leu	Glu	Arg	Glu	Glu	Arg
Gly				260				265					270		

(2) SEQ ID NO: 4 的资料

(i) 序列特征:

(A) 长度: 524 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 4

```
Met Thr Pro Ala Asp Lys Pro Thr Asp Glu Arg Ser Pro Val Leu Asp
1          5          10          15

Leu Ser Gly Val Ser Val Ala Tyr Gly Thr Arg Thr Val Leu His Gly
20          25          30

Ile Asp Leu Arg Leu Ala Pro Gly Gln Val Leu Ala Val Leu Gly Ala
35          40          45

Ser Gly Ser Gly Lys Ser Thr Leu Ala Gln Ala Ala Leu Gly Leu Leu
50          55          60

Pro Pro Gly Gly Arg Val Thr Ala Gly Arg Val Thr Val Ala Gly His
65          70          75          80

Asp Ile Thr Ala Leu Ala Pro His Arg Leu Arg Ala Leu Arg Gly Thr
85          90          95

Val Thr Gly Leu Val Pro Gln Asp Gln Ala Val Ser Leu Asp Pro Leu
100         105         110

Val Arg Val Gly Ala Gln Val Thr Glu Thr Leu Arg Ala His Arg Leu
115         120         125

Glu Asp Arg Arg Glu Ala Ala Arg Arg Ala Val Pro Leu Leu Gly Glu
130         135         140

Ala Gly Ile Glu Ala Pro Gly Pro Leu Ala Arg Ala Tyr Pro His Ala
145         150         155         160

Leu Ser Gly Gly Gln Arg Gln Arg Val Leu Val Ala Gly Ala Phe Ala
165         170         175

Ala Arg Pro Pro Leu Val Val Ala Asp Glu Pro Thr Ser Ala Leu Asp
180         185         190

Ala Thr Val Arg Arg Arg Val Met Asp Arg Phe Ala Ala Leu Val Ala
195         200         205

Ala His Gly Thr Ala Val Leu Leu Val Thr His Asp Phe Arg Leu Ala
210         215         220
```

Arg Glu Arg Ala Asp Gln Val Ala Val Leu Ala Asp Gly Arg Leu Val
 225 230 235 240
 Glu Ser Gly Pro Ala Ala Arg Val Leu Asp Arg Pro Ala His Pro Tyr
 245 250 255
 Thr Arg Arg Leu Thr Gly Ala Gly Arg Arg Val Ala Ala Arg Gly Thr
 260 265 270
 Ala Pro Arg Ala Ser Gly Thr Pro Val Val Arg Ala Arg Asp Leu Val
 275 280 285
 Lys Glu Tyr Arg Arg Asp Gly Arg Arg Val Arg Ala Val Asp Gly Val
 290 295 300
 Gly Phe Thr Val Arg Glu Gly Glu Phe Phe Ala Leu Val Gly Glu Ser
 305 310 315 320
 Gly Ser Gly Lys Ser Thr Thr Ala Arg Leu Val Thr Gly Leu Thr Ala
 325 330 335
 Pro Thr Ser Gly Ala Val Glu His Ala Pro Ala Pro Val Arg Pro Gln
 340 345 350
 Leu Val Gln Gln Ser Pro Tyr Ala Ala Phe Asp Pro Arg Trp Thr Val
 355 360 365
 Arg Arg Ile Val Glu Glu Pro Leu Arg Ala Arg His Val Pro Gly Ala
 370 375 380
 Arg Arg Arg Ala Arg Leu Arg Glu Leu Leu Ala Leu Val Gly Leu Asp
 385 390 395 400
 Glu Glu Leu Leu Ala Arg Arg Pro Arg Glu Leu Ser Gly Gly Gln Arg
 405 410 415
 Gln Arg Val Ala Leu Ala Arg Ala Leu Ala Pro Glu Pro Arg Leu Leu
 420 425 430
 Val Cys Gly Glu Pro Val Ala Ala Leu Asp Pro Val Ala Arg Glu Arg
 435 440 445
 Val Val His Leu Leu Glu Arg Leu Arg Ala Glu Leu Gly Leu Thr Cys
 450 455 460
 Leu Phe Val Ser His Glu Leu Asp Val Val Arg Arg Leu Cys Gly Arg
 465 470 475 480
 Val Ala Val Met Arg Gly Gly Arg Leu Leu Glu Ser Gly Pro Val Gly
 485 490 495
 Glu Val Leu Ser Ala Pro Ser Gln Pro Tyr Thr Arg Ala Leu Leu Ala
 500 505 510
 Ala Glu Ala Gly Pro Ser Asp Thr Pro Gly Ala Gly
 515 520

(2) SEQ ID NO: 5 的资料

(i) 序列特征:

(A) 长度: 216 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 5

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

(2) SEQ ID NO: 6 的资料

(i) 序列特征:

(A) 长度: 517 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 6

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

Leu Leu Ala Leu Val Leu Gly Ile Val Gln Ser Gly Pro His Gly Trp
 225 230 235 240
 Gly Ser Ala Ala Thr Leu Val Pro Leu Leu Gly Gly Leu Ala Leu Leu
 245 250 255
 Gly Ala Phe Val Leu Val Glu Gly Arg Phe Ala Pro Gln Pro Leu Ile
 260 265 270
 Pro Leu Gly Ile Phe Arg Ser Arg Ser Val Val Ala Ala Asn Val Val
 275 280 285
 Ala Met Thr Ser Gly Ala Ala Leu Phe Ser Met Phe Tyr Phe Leu Thr
 290 295 300
 Leu Phe Leu Asn Gln Val Arg Asp Tyr Ser Pro Leu Arg Thr Gly Phe
 305 310 315 320
 Ala Tyr Leu Pro Leu Ala Leu Ala Ile Met Val Ala Ala Gln Phe Ser
 325 330 335
 Ala Ala Leu Val Arg Val Leu Gly Pro Arg Thr Thr Leu Leu Val Ser
 340 345 350
 Met Ala Leu Thr Ala Ala Gly Leu Leu Trp Leu Ser Arg Leu Thr Glu
 355 360 365
 Asp Ser Gly Phe Ala Gly Gly Leu Leu Gly Pro Thr Leu Val Val Gly
 370 375 380
 Ile Gly Gln Gly Ile Ser Met Ser Ala Ser Ala Ile Ala Gly Val Ala
 385 390 395 400
 Gly Val Arg Pro Gln Gln Ala Gly Leu Ala Ser Gly Leu Leu Asn Ala
 405 410 415
 Thr Arg Gln Leu Gly Gly Ala Leu Gly Leu Ala Val Val Ala Ala Val
 420 425 430
 Ala Thr Ser Arg Ala Asp Gly Leu Leu Asp Gly Val Ala Arg Pro Thr
 435 440 445
 Ala Glu Leu Ala Arg His Ala Gln Ala Ser Gly His Pro Leu Ser Ile
 450 455 460
 Ala Val Ala Ala Ala Leu Ser Ala Val Gly Leu Leu Ala Ser Leu Ala
 465 470 475 480
 Ala Pro Gly Arg Ser Pro Ala Pro Thr Gly Thr Arg Thr Gly Gly Asp
 485 490 495
 Ser Ala Ala Pro Ala Pro Ala Ala Ala Pro Ala Ala Thr Gly Thr Thr
 500 505 510
 Gly Pro Gly Glu Ile
 515

(2) SEQ ID NO: 7 的资料

(i) 序列特征:

(A) 长度: 282 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 7

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

(2) SEQ ID NO: 8 的资料

(i) 序列特征:

(A) 长度: 171 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 8

```
Met Asn Glu Thr Arg Thr Ala Ala Arg Thr Gly Gln Val Gly Pro Val
1          5          10          15
Asp Ala Glu Gly Phe Arg Ala Ala Met Ser Cys Phe Pro Ala Gly Val
20          25          30
Val Leu Val Thr Thr Arg Glu Glu Asp Gly Thr Pro Arg Gly Phe Thr
35          40          45
Ala Ser Ser Phe Cys Ser Val Ser Leu Ala Pro Pro Leu Val Ser Val
50          55          60
Cys Gln Gly Thr Gly Ala Gln Ser Tyr Gly Ala Phe Gln Glu Cys Ala
65          70          75          80
Arg Phe Ala Val Ser Val Leu Arg Ser Gly Gln Arg Glu Leu Ala Ser
85          90          95
Arg Phe Ala Thr Arg Gly Ala Asp Lys Phe Gly Gly Gly Gly Leu Val
100         105         110
Ala Leu Glu Gly Ser Gly Leu Leu Val Ala Ala Asp Ala Leu Val Thr
115         120         125
Leu Glu Cys Ala Val His Ala Arg His Leu Ala Gly Asp His Val Ile
130         135         140
Leu Val Gly Glu Val Arg Gly Val Gly Gln Gly Glu Gly Glu Pro Leu
145         150         155         160
Val His Trp Glu Arg Gly Phe Arg Ala Leu Arg
165         170
```

(2) SEQ ID NO: 9 的资料

(i) 序列特征:

(A) 长度: 336 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 9

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

Asp	Gly	Arg	Arg	Val	Phe	Arg	Trp	Ala	Met	Asp	Glu	Val	Thr	Pro	Arg	
225					230					235					240	
Ala	Ala	Glu	Leu	Leu	Arg	Glu	Ser	Gly	Ile	Glu	Pro	Glu	Ala	Leu	Asp	
				245					250						255	
Ala	Phe	Val	Pro	His	Gln	Ala	Asn	Leu	Arg	Met	Ile	Glu	Leu	Met	Ala	
			260					265						270		
Glu	Arg	Leu	Gly	Leu	Pro	Glu	Arg	Thr	Ala	Val	Ala	Arg	Asp	Val	Val	
		275					280						285			
Arg	Ala	Gly	Asn	Thr	Ser	Ala	Ala	Ser	Val	Pro	Leu	Ala	Leu	Glu	Ala	
	290					295					300					
Leu	Leu	Asp	Ser	Gly	Glu	Val	Gly	Ser	Gly	Asp	Arg	Ala	Leu	Leu	Val	
305					310					315					320	
Gly	Phe	Gly	Ala	Gly	Leu	Asn	Tyr	Ala	Ala	Gln	Val	Val	Glu	Leu	Pro	
				325					330						335	

(2) SEQ ID NO: 10 的资料

(i) 序列特征:

(A) 长度: 85 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 10

```
Met Ser Ala Phe Thr Leu Thr Glu Phe Lys Lys Leu Val Glu Gln Ser
1           5           10           15
Tyr Asp Ala Glu Ser Ala Glu Ala Leu His Gly Gln Ala Leu Asp Thr
20           25           30
Ser Phe Thr Asp Leu Gly Tyr Asp Ser Leu Thr Val Tyr Glu Ile Val
35           40           45
Thr Arg Ile Gln Asp Glu His Gly Val Thr Val Pro Asp Glu Glu Leu
50           55           60
Asp Leu Leu Asp Thr Pro Arg Ala Leu Ile Ala Tyr Val Asp Ala Arg
65           70           75           80
Ala Gly Ser Arg Thr
85
```

(2) SEQ ID NO: 11 的资料

(i) 序列特征:

(A) 长度: 351 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 11

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

(2) SEQ ID NO: 12 的资料

(i) 序列特征:

(A) 长度: 425 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 12

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

Ser Arg Pro Phe Asp Arg Glu Arg Asn Gly Phe Val Leu Ala Glu Gly
 225 230 235 240
 Ala Ala Val Leu Val Leu Glu Glu Leu Gly His Ala Arg Ala Arg Asp
 245 250 255
 Ala His Val Tyr Ala Leu Val Ser Gly Tyr Ala Thr Arg Cys Asn Ala
 260 265 270
 Tyr His Met Thr Gly Leu Thr Pro His Gly Arg Glu Met Ala Glu Ala
 275 280 285
 Ile Arg His Ala Leu Ala Glu Ser Gly Thr Asp Pro Ala Ala Val Asp
 290 295 300
 Tyr Val Asn Ala His Gly Ser Gly Thr Lys Gln Asn Asp Arg His Glu
 305 310 315 320
 Thr Ala Ala Phe Lys Ala Thr Leu Gly Glu Arg Ala Arg Ser Val Pro
 325 330 335
 Val Ser Ser Ile Lys Ser Met Ile Gly His Ser Leu Gly Ala Ile Gly
 340 345 350
 Ser Leu Glu Ile Ala Ala Trp Ala Leu Ala Met Glu Tyr Gly Val Val
 355 360 365
 Pro Pro Thr Ala Asn Leu Asp Thr Pro Asp Pro Glu Cys Asp Leu Asp
 370 375 380
 Tyr Val Pro His Glu Ala Arg Glu Thr Arg Val Asp Arg Val Leu Ser
 385 390 395 400
 Val Gly Ser Gly Phe Gly Gly Phe Gln Ser Ala Met Val Leu Thr Arg
 405 410 415
 Asp Thr Gly Ala Arg Leu Pro Thr Ala
 420 425

(2) SEQ ID NO: 13 的资料

(i) 序列特征:

(A) 长度: 426 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 13

```
Met Thr Thr Ala Pro Ser Arg Thr Ala Gln Gly Ala Pro Pro Gly Ala
1          5          10          15

Ala Leu Pro Pro Val Phe Thr Gly Ile Gly Val Ala Ala Pro Asn Gly
20          25          30

Leu Gly Thr Glu Glu Trp Trp Ala Ala Thr Leu Arg Gly Glu His Gly
35          40          45

Leu Arg Pro Val Thr Glu Tyr Asp Ala Ser Gly His Pro Gly Gly Leu
50          55          60

Val Gly Arg Val Pro Asp Phe Asp Ala Ala Arg His Leu Pro Gly Arg
65          70          75          80

Leu Leu Pro Gln Thr Asp Arg Val Thr Arg Leu Ala Leu Val Ala Ala
85          90          95

Asp Glu Ala Leu Lys Asp Ala Ala Val Asp Pro Ala Arg Leu Pro Glu
100         105         110

Tyr Gly Ala Ser Ala Val Thr Ser Asn Ala Thr Gly Gly Phe Glu Phe
115         120         125

Thr His Arg Glu Ile Arg Lys Leu Trp Thr Glu Gly Pro Ala Arg Val
130         135         140

Ser Val Tyr Glu Ser Phe Ala Trp Phe Tyr Ala Val Asn Thr Gly Gln
145         150         155         160

Ile Ser Ile Arg His Gly Met Arg Gly Pro Gly Ala Val Val Val Ala
165         170         175

Asp Gln Ala Gly Gly Leu Asp Ala Leu Gly Gln Ala Arg Arg Val Leu
180         185         190

Arg Lys Gly Gly Val Leu Ala Val Ser Gly Gly Val Glu Ser Ala Leu
195         200         205

Asp Pro Trp Gly Leu Ala Ala His Ala Ser Ser Gly Thr Leu Ser Arg
210         215         220
```

Ser Gly Asp Pro Ala Thr Ala Tyr Leu Pro Phe Asp Arg Arg Ala Leu
 225 230 235 240
 Gly Thr Val Val Gly Glu Gly Gly Ala Leu Leu Thr Leu Glu Thr Pro
 245 250 255
 Arg His Ala Glu Glu Arg Asp Ala Pro Arg Ile Tyr Gly Glu Leu Ala
 260 265 270
 Gly Tyr Ala Ala Thr Phe Asp Pro Pro Ala Gly Ser Gly Arg Pro Pro
 275 280 285
 Gly Leu Glu Arg Ala Ala Arg Leu Ala Leu Ala Asp Ala Gly Leu Ala
 290 295 300
 Pro Gly Asp Val Asp Val Val Phe Ala Asp Ala Ala Gly Leu Pro Ala
 305 310 315 320
 Ala Asp Ala Ala Glu Ala Ala Ala Leu Arg Ala Leu Phe Gly Pro Gly
 325 330 335
 Gly Val Pro Val Ser Val Pro Lys Thr Gln Thr Gly Arg Leu Ala Ser
 340 345 350
 Gly Gly Pro Ala Leu Asp Val Ala Ala Ala Leu Leu Ala Leu Arg Asp
 355 360 365
 Gly Leu Val Pro Pro Ala Val His Leu Asp Glu Val Asp Pro Ala Tyr
 370 375 380
 Gly Leu Asp Leu Val Arg Asp Thr Pro Arg Ala Leu Pro Leu Arg Thr
 385 390 395 400
 Ala Leu Val Leu Ala Arg Gly His Gly Gly Phe Asn Ala Ala Val Val
 405 410 415
 Val Arg Gly Arg Arg Arg Pro Arg Thr Ala
 420 425

(2) SEQ ID NO: 14 的资料

(i) 序列特征:

(A) 长度: 83 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 14

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

(2) SEQ ID NO: 15 的资料

(i) 序列特征:

(A) 长度: 160 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 15

Met	Thr	Thr	Thr	Thr	Pro	Pro	Asp	Asp	Val	Arg	Ala	Gly	Ser	Leu	Pro	1	5	10	15
Gly	Asp	Ala	Ala	Arg	Ser	Ala	Ala	Leu	Tyr	Thr	Glu	Val	Gln	Ala	Phe	20	25	30	
Tyr	Ala	Arg	Gln	Ala	His	His	Leu	Asp	Ala	Val	Arg	Ala	Glu	Glu	Phe	35	40	45	
Ala	Ala	Thr	Phe	Ala	Ala	Glu	Gly	Val	Phe	Ala	His	Ser	Pro	Asp	Thr	50	55	60	
Pro	Ala	Ala	Arg	Gly	Arg	Ala	Ala	Ile	Ala	Glu	Glu	Val	Arg	Gly	Phe	65	70	75	80
Asn	Ala	Arg	Arg	Phe	Ala	Asp	Asp	Pro	Val	Gln	Arg	Arg	His	Trp	Phe	85	90	95	
Ser	Met	Leu	Asp	Val	Arg	Pro	Gly	Glu	Asp	Gly	Ala	Val	Glu	Thr	Glu	100	105	110	
Phe	Tyr	Ala	Leu	Val	Val	Val	Thr	Arg	Pro	Asp	Ala	Ala	Leu	Pro	Val	115	120	125	
Ile	Gly	Pro	Ser	Cys	Val	Val	Arg	Asp	Val	Leu	Val	Arg	Glu	Gly	Gly	130	135	140	
Glu	Leu	Arg	Thr	Leu	Ser	Arg	Gln	Val	Thr	Gln	Asp	Arg	Thr	Leu	Leu	145	150	155	160

(2) SEQ ID NO: 16 的资料

(i) 序列特征:

(A) 长度: 274 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 16

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

(2) SEQ ID NO: 17 的资料

(i) 序列特征:

(A) 长度: 331 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 17

```
Met Thr Thr Thr Thr Thr Gly His Gln Arg Pro Gly Ser Ala Glu His
1           5           10           15
Ser Ala Arg Leu Ala Ala Pro Pro Ala Ser Ala Tyr Glu Leu Val Ala
20           25           30
Asp Val Thr Arg Trp Pro Leu Leu Phe Thr Pro Cys Leu His Ala Glu
35           40           45
Val Leu Glu Ser Gly Pro Gly Thr Glu Arg Val Arg Leu Trp Ala Leu
50           55           60
Thr Gly Glu Gln Val Arg Gly Trp Thr Ser Arg Arg Thr Leu Asp Ser
65           70           75           80
Glu Gly Leu Arg Val Gly Phe Arg Gln Glu Asp Ser Ala Pro Pro Leu
85           90           95
Ala Ala Met Gly Gly Glu Trp Arg Phe Thr Glu Glu Gly Glu Asp Thr
100          105          110
Arg Ala Val Leu Ala His Asp Trp Thr Leu Thr Glu Pro Gly Ala Ala
115          120          125
Pro His Arg Trp Val Thr Glu Thr Leu Asp Arg Asn Ser Thr Ala Glu
130          135          140
Ile Gly Ala Val Thr Ala Trp Ala Ala Arg Thr His Ala Ala Gly Gly
145          150          155          160
Ala Asp Ala Leu Leu Phe Ser Phe Thr Asp Ser Leu Asp Ile Ala Ala
165          170          175
Pro Ala Pro Asp Val Tyr Ala Phe Leu Asp Ala Ala Asp Gln Trp Pro
180          185          190
Ala Arg Leu Pro His Val Ser Arg Val Ala Phe Ser Thr Thr Pro Ala
195          200          205
```

Thr	Pro	Leu	Thr	Ala	Gly	Ala	Glu	Val	Gln	His	Leu	Glu	Met	Glu	Thr	210	215	220	
Arg	Ala	Asp	Asp	Gly	Thr	Arg	His	Leu	Thr	Arg	Ser	Ile	Arg	Leu	Gly	225	230	235	240
Phe	Ala	Gly	Arg	Leu	Leu	Val	Tyr	Lys	Gln	Thr	Thr	Leu	Pro	Ala	Pro	245	250	255	
Leu	Leu	Gly	His	Ala	Gly	Ser	Trp	Ala	Leu	Glu	Pro	Leu	Pro	Gly	Gly	260	265	270	
Gly	Thr	Arg	Val	Thr	Ala	Arg	His	Arg	Val	Ala	Leu	Asp	Pro	Asp	Ala	275	280	285	
Val	Thr	Glu	Arg	Phe	Gly	Ala	Gly	Thr	Thr	Leu	Ala	Ala	Ala	Arg	Asp	290	295	300	
Thr	Val	Arg	Ala	Leu	Leu	Gly	Gly	Asn	Ser	Arg	Arg	Thr	Leu	Glu	Ala	305	310	315	320
Ala	Arg	Ala	His	Thr	Glu	Ser	Ala	Gly	Glu	Arg						325	330		

(2) SEQ ID NO: 18 的资料

(i) 序列特征:

(A) 长度: 314 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 18

```
Met Thr Thr Thr Gly Pro Ala Thr Pro Leu Ala Pro Gly Pro Ala Ser
1      5      10      15
Ala Glu Thr Val Ala Leu Ala Asp Gly Val His Ser Trp Leu Gln Pro
      20      25      30
Asp Gly Gly Trp Cys Val Ser Asn Ala Gly Ile Leu Leu Ala Pro Asp
      35      40      45
Arg Val Ala Leu Val Asp Thr Ala Ala Thr Glu Ala Arg Ser Arg Ala
      50      55      60
Leu Gly Ala Ala Val Ala Gly Leu Ser Pro His Pro Val Arg Leu Leu
65      70      75      80
Val Asn Thr His Phe His Gly Asp His Ser Phe Gly Asn Gly Ile Leu
      85      90      95
Gly Lys Asp Ala Val Ile Val Ala His Glu Arg Thr Arg Thr Glu Met
      100      105      110
Ala Glu Ala Gly Leu Gly Leu Thr Gly Leu Trp Pro Gly Val Asp Trp
      115      120      125
Gly His Val Asp Pro Val Leu Pro Gln Leu Thr Tyr Arg Arg Arg Leu
130      135      140
Thr Leu His His Gly Asp Leu Arg Val Glu Leu Ile His Pro Gly Pro
145      150      155      160
Ala His Thr Thr Asn Asp Thr Leu Val Trp Leu Pro Glu Gln Arg Val
      165      170      175
Leu Phe Ala Gly Asp Val Leu Leu Pro Gly Ala Thr Pro Phe Val Leu
      180      185      190
Met Gly Ser Val Thr Gly Ser Leu His Thr Leu Arg Leu Leu Arg Arg
195      200      205
```

Leu	Gly	Pro	Arg	Val	Val	Val	Gly	Gly	His	Gly	Pro	Leu	Ala	Gly	Pro	210	215	220
Glu	Val	Ile	Glu	Glu	Thr	Glu	Arg	Tyr	Leu	Leu	Arg	Leu	Arg	Arg	Ile	225	230	235 240
Ala	Thr	Glu	Gly	His	Ala	Ala	Gly	Leu	Thr	Pro	Leu	Glu	Ala	Ala	Arg	245	250	255
Arg	His	Gly	Pro	Gly	Pro	Phe	Ala	His	Trp	Ser	Glu	Pro	Glu	Arg	Leu	260	265	270
Ala	Ala	Asn	Leu	His	Arg	Ala	Tyr	Ala	Glu	Leu	Gly	Pro	Ala	Pro	Leu	275	280	285
Gly	Thr	Pro	Leu	Asp	Val	Leu	Ala	Cys	Phe	Gly	Asp	Leu	Ile	Ala	Tyr	290	295	300
Asn	Asp	Gly	Glu	Leu	Pro	Val	Ser	His	Ala							305	310	

(2) SEQ ID NO: 19 的资料

(i) 序列特征:

(A) 长度: 305 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 19

```
Met Thr Gly Arg Val Thr Gly Thr Met Ala Gly Pro Leu Pro Gly Thr
1          5          10          15
Val Leu Val Ala Gly Ala Ser Gly Ala Val Gly Gly Arg Leu Val Ser
20          25          30
Arg Leu Leu Asp Ser Gly Val Pro Val Arg Ala Leu Val Arg Ser Ala
35          40          45
Ala Arg Gly Arg Ala Leu Ala Ala Val Gly Ala Glu Thr Val Val Gly
50          55          60
Asp Leu Ala Asp Thr Ala Ser Leu Ala Ala Ala Leu His Gly Val Glu
65          70          75          80
Arg Ala Phe Leu Leu Leu Gln Asp Glu Ala Gly Ala Pro Phe Ala Arg
85          90          95
Ala Ala Ala Lys Ala Arg Gly Leu Arg Glu Val Val Val Leu Ser Ala
100         105         110
Thr Ala Ala Ala His Pro Glu Tyr Asp Asn Pro Met Phe Gln Lys His
115         120         125
Val Arg Gly Glu Arg His Val Arg Gly Ser Gly Val Pro Trp Val Phe
130         135         140
Leu Arg Pro Gly Ala Phe Ala Ser Leu Ala Leu Leu Trp Ala Pro Ala
145         150         155         160
Leu His Gly Asp Gly Val Val Arg Val Pro His Ala Gly Leu Ala Val
165         170         175
Pro Leu Ile Asp Pro Arg Asp Ile Ala Asp Val Ala Ala Ala Val Leu
180         185         190
```

Thr Ala Pro Val Glu Arg Trp Ala Gly Lys Ala Val Pro Leu Ser Gly
 195 200 205
 Pro Glu Val Leu Thr Leu Pro Gly Arg Thr Ala Val Leu Ala Ala Glu
 210 215 220
 Leu Gly Arg Pro Leu Arg Val Glu Pro Val Pro Glu Glu Glu Trp Val
 225 230 235 240
 Gln Leu Thr Ala Arg His Ile Pro Glu Pro Tyr Ala Arg Ala Leu Ala
 245 250 255
 Gly Val Glu Arg Phe Phe Thr Glu His Pro Pro Ala Val Ser Pro Gly
 260 265 270
 Val Thr Asp Val Thr Gly Arg Pro Ala Arg Ser Phe Ala Thr Trp Val
 275 280 285
 Arg Asp His Ala Ala Ser Phe Ala Pro Gly Pro Pro Thr Val Val Gly
 290 295 300
 Gly
 305

(2) SEQ ID NO: 20 的资料

(i) 序列特征:

(A) 长度: 396 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 20

```
Met Thr Glu Gln Ser Thr Thr Leu Ala Ala Pro Ala His Ala Glu Arg
1          5          10          15
Thr Gly Arg Arg Thr Lys Gly Pro Gly Arg Pro Pro Asp Arg Ser Arg
20          25          30
Asn Gly Pro Gly Arg Ser Pro Arg Ser Arg Pro Gly Thr Arg Ala Ala
35          40          45
Val Arg Arg Ser Gly Val Ser Pro Gly Lys Ser Ser Ala Gly Ile Val
50          55          60
Ala Ala Gly Phe Ala Arg His Phe Val Pro Ala Ala His Gly Gly Ala
65          70          75          80
Glu Gly Gly Phe Gly Glu Leu Ala Glu Ala Val Leu Arg Leu Gly Thr
85          90          95
Gly Cys Thr Ser Ala Ala Trp Ala Ala Ser Leu Ser Ala Tyr Ala Gly
100         105         110
Arg Tyr Ala Ala Tyr Leu Pro Glu Glu Gly Gln Ala Glu Val Trp Ala
115         120         125
Glu Gly Pro Asp Ala Leu Leu Ala Gly Ala Leu Val Pro Ser Gly Thr
130         135         140
Val Thr Pro Val Pro Gly Gly Trp Arg Leu Asp Gly Ala Trp Pro Tyr
145         150         155         160
Ile Ser Gly Val Arg His Ala Ala Trp Val Leu Ala Cys Ala Thr Val
165         170         175
Pro Gly Gly Glu Gly Glu Glu Gly Pro Glu Val Arg Phe Phe Ala Gly
180         185         190
Pro Arg Ala Ala Pro Arg Val Glu Arg Thr Trp Asn Thr Thr Gly Met
195         200         205
```

Arg Ala Thr Gly Ser Asp Thr Leu Val Leu Asp Asp Val Leu Val Pro
 210 215 220
 Ala His Arg Ser Phe Pro Arg Thr Arg Val Leu Ala Gly Gln Arg Pro
 225 230 235 240
 Arg Val Ala Gly Ala Val Pro Thr Val Arg Met Ala Arg Val Gly Ala
 245 250 255
 Leu Pro Val Val Thr Pro Leu Val Gly Ala Ala Arg Gly Ala Leu Arg
 260 265 270
 Ala Trp Thr Glu Arg Ala Ala Gln Gly Arg Ala Pro Ser Pro Gly Ala
 275 280 285
 Leu Gly Glu Leu Ser Arg Ala Ala Gly Glu Ala Asp Ala Ala Glu Leu
 290 295 300
 Leu Val Leu Arg Ala Ala Ala Ala Asp Gly Thr Val Ser Leu Pro
 305 310 315 320
 Glu Pro Ala Ala Ala Val Arg Gly Lys Arg Asp Thr Ala Leu Ala Ala
 325 330 335
 Glu Leu Ala Val Gly Ala Val Gln Arg Leu Val Arg Ala Ser Gly Thr
 340 345 350
 Ser Gly Gln Ser Pro His Asp Pro Val Gln Arg Phe Trp Arg Asp Val
 355 360 365
 Gln Thr Gly Ala Ser His Val Ala Leu Ser Pro Glu Ala Ala Gly Ala
 370 375 380
 Ala Tyr Gly Ala Trp Ala Val Gln Glu Ala Asn Arg
 385 390 395

(2) SEQ ID NO: 21 的资料

(i) 序列特征:

(A) 长度: 108 个氨基酸

(B) 类型: 氨基酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii) 分子类型: 蛋白质

(iii) 假设的: 不是

(iv) 反意义: 不是

(ix) 序列描述: SEQ ID NO: 21

Met	Ser	Gly	Gly	His	Met	Val	Thr	Met	Ile	Asn	Gln	Met	Leu	Leu	
1				5					10					15	
His	Gly	Asp	Glu	Ser	Arg	Phe	Leu	Ala	Val	Leu	Glu	Glu	Ile	Cys	
				20					25					30	
Ala	His	Met	Arg	Ala	Gln	Pro	Gly	Phe	Leu	Ser	Leu	Arg	Leu	His	
				35					40					45	
Arg	Ser	Pro	Asp	His	Pro	Glu	Arg	Trp	Ala	Met	Leu	Ala	Asp	Trp	
				50					55					60	
Ser	Asp	Ala	Ala	Ala	His	Arg	Ala	Ala	Ala	Ser	Ala	Pro	Gly	Ile	
				65					70					75	
Arg	Pro	Ala	Phe	Ala	Arg	Leu	Arg	Ala	Glu	Ala	His	Thr	Ala	Pro	
				80					85					90	
Gln	Val	Tyr	Ala	Pro	Val	Pro	Thr	Pro	Gly	Ala	Pro	Ala	Gly	Asp	
				95					100					105	
Ser	Leu	Ala													

(2) SEQ ID NO: 22 的资料

(i) 序列特征:

(A) 长度: 24379 个碱基对

(B) 类型: 核酸

(C) 链数: 单链

(D) 几何结构: 线性

(ii)分子类型: 蛋白质

(iii)假设的: 不是

(iv)反意义: 不是

(ix)序列描述: SEQ ID NO: 22

CAATTGCCAT GGAATGCCCC GTTATAAGGG AAGTGAATG TGTCACCGCC GTCGTCGCGT
60

GGGGCGGTGG GCGGCTAATG TGAGCGATCC CGGTGGCCGC GACCGTACTC ATCGCGGCGC
120

GTCACCGATT GACCCGAGTG CCTCCCGCCG CGTCTCCACC CCACCTGGAG GGCGGGTCGG
 180

CGCCGCGTGC CCCGTGACGG AGGATTTGGA CAACCGTGGA CGCGCAAGGC CGTGGGCGGG
 240

GGGAAGAATT CGACCCCGCA GACCAATGGG TACTCAATCC GCAGACCGGC GAATACGAAC
 300

TGCGACTCGG CGGCTCCGCT GAGCAGTCGC AGGAGCCGCC GGGGTCCGGC GGCGGGGCCT
 360

GGGACCACGT CGACGGGTCC GCGCGTCCCC GCAGAACCGA GCCGTACCGC TCCGAGCCCC
 420

GGGTGTCCGA GGCAGCGGCG GCTCCCCGCG CACGGAGAAC GCCCCGCGCG CATGGCGGTC
 480

GCGCGGCGGC GCAGTAGGAG GCGGCGGCCG GCGGTTCGTCC CTCGGCCCGC AAACAGGCTT
 540

CCGGGCGGGC CGCCTCGCGC AAGGCGCGCA AGGGCGGCGG CAAGAAGAAG GCGCTGCTGT
 600

GGACGGCCGG GGGGCTCGGT TTCGCGCTCG TCGCGGTGGC CGGCAGCGGC TACGCGTACT
 660

ACCAGCACCT CAGCGGCAAC ATCGACAAGA TTGACGTCGG CGACGCGGGC AACAAGGACG
 720

CCGCCCCGA CGGGCCGATC AACATCCTGA TCATCGGCAC CGACAAGCGC ACCGGCAAGG
 780

GCAACGAGGG GTGCGGGGGC AAGGACAGCC CTGGCCACGC GGACACCAAC ATCTTGCTGC
 840

GCGTCTCCGC CGACCGCACC AACACGACCG GCCTGAGCAT TCCGCGTGAC CTCATCACGA
 900

ACATCCCGGA CTGCCTGACC ACCCAGGACG ACGGTTCGAA GAAGACGATC CCCGGCACGC
 960

AGAACGTCCG CTTCAACACG AGCCTCGGGC AGGAGGGCCG CGACCCCGGC TGCACGATGC
 1020

GCACGGTCAC CGAGCTGACC GGGCTCAAGG TCGACCACTT CATGATGGCC GACTTCAACG
 1080

CGGTCAAGAA CCTGACGACG GCCGTGAACG GCGTCGAGGT GTGCGTGGCG AAGGACGTCTG
 1140

ACGACCCCGA CTCGCACCTC AAGCTCTCGG CGGGGACGCA CAAGGTGCAG GGCGAGCAGG
 1200

CGCTCGCCTT CGTCCGCACC CGGCACAGCT TCGGCAACCA GGGCGACCTC GACCGCATCA
 1260

AGGTCCAGCA GCAGTTCCTC GGTTCGCTCG CGCGGCAGCT CAAGTCCGAG GACACGCTGA
 1320

CGAGTCCGAA GAAGCTCTAC AAGGTCGCCG AGGCGGCGAC CGACGCGCTG ACCGTGGACT
 1380

CGGGGATCGG CAGCATCACG AAAGTATGT CGCTGGCCAA GGAGTTGCAG CACATCAATC
 1440

CGAAGAACAT CACCTTCGTG ACGCTCCCCG TGGTCGACAA CCCGGCCGAG AAGGTCAAGG
 1500

CGACGGTCGT CCTCAATGAG ACGGACGCTG ATCCGCAGCA GTCGGCTCTC GGGCAGAGTC
 1560

TCGACGTGGG CCGCCAGCTC GTCGACTCGC TCACCGACCA GGACCCGCGC GACGGCAAGA
 1620

CCGTCCCGTG GCTCGCGACG CGCTGGAAGG CCGACCCCGA GGCCACCCGC TTCACCTTCA
 1680

CCCTGCGCGC GGGCGCCACT TTCAGCGACG GCACACCGGT CGACGCGCGC GCGGTCAAGG
 1740

CCAACTTCGA CGCCGTGCAC GAGCTCGGCG CCGCCGCCTC GCGGGGGGCC GTGTACCTCG
 1800

ACGGCTACCG CGAGACGCGC GTCAGCGGCG CCCGTACGCT CACGGTCGTC TTCGACAAGC
 1860

CGAACGCCCA GTTCTGCGC GGCACGTCCA CCGTCTCACT GGGCCTGCTC TCGCCGGGCA
 1920

GCCTGCGCCG CACCCGCGAG GAGCGCTGCA CCGGGCGCCT CGTCGGCTCC GGTCCCTTCG
 1980

TGCTCGACCG CTACCGGCCG AACACCTCCG TCACGCTGGA CCGGAGAAAG GGCTACTCCT
 2040

GGGGCTCGCG CCTGTGGCAA CGGGAGGGCG GCGCGTACCT GGAGGGGGTC GAGTACCGGA
 2100

TCGTGCCCGA GAACACGACC CGCTCGGGCG CCCTGTCCGC CGGTCAGCTC GACGTCGCCA
 2160

CCGCGCTCGC GCCGCAGGAC AGGGAGCGGT TCTCCGCCCC CGGCTGGTCG CTGCTGACCC
 2220

GTACCGCGCC CGGCGTCGAC CTCAGCCTGT ACGTCAACGC GCGCCGCACG GCGCTGCGCG
 2280

AGGCGGCGGT GCGCCAGGCC CTCCAGAAGG GCATCGACCG CGAAGCCGTC GCCACGACCT
 2340

TCCTCAGTTC CCGCAAGCTC GCGGCGACCA GCGTGCTCTC CTCCACGACC CCCGGCTACA
 2400

CCGACCTCGG CGACCGCCTC GCCCACGACC CTGCGGGTGC GAGGCGGCTG CTCGACAAGG
 2460

CCGGATGGCG GCCCGGGGCG GACGGCATCA GGGTCAAGAA CGGCGTCAGG CTGCGACTCG
 2520

ACGCCGTCTT CGTCCGGCAG CAGAGCCTCG AACTCGTGCA GCAGCAGCTC AAGGACATCG
 2580

GCGTCGAACT GCGGCTGAGA CAGCTCACCG TCTCCCGTTT CCCCGAAGTG CTCGCCGCGG
 2640

GCAGCTACGA CCTGAGCCTC CAGAGCGCCA ACCGCGCCGA TCCCGACGTC CTCACCACGG
 2700

CCTTCGCGGG CGGCACCCCC GTCGCCGACG CGCGCCTGCG CTCCGAACTG CGCCGGGCCA
 2760

CCTCCTCCAC CGACGAGGCG ACGCGCTCCT CGCTCTTCGC CGCGGCGCAA CGGCGTCTCG
 2820

TCGACGAGGG GCACGTCCTG CCGCTCAACG AGACCGAGGA GACCGCCGCG CTCTCGACGC
 2880

GGGTGCACGG GCTGACGCGG GACGCCTCCA ACCGGCTGGT CCTGCACGAC ACCTGGACCA
 2940

CCGGGTGACG GCCCGCTACC TGGCCCGCCG GCTGGGCCGC GTCGTCCTCG TCGTCTGGGC
 3000

CGCGTACACC CTGTCCTTCG CCGTCCTCTA CCTGCTGCCG GGAGATCCCG TGCAGACCAT
 3060

GCTGAGCGGC GCGGCGGGCG GGGACGGGGC GGCCGTCGAC CCGCACGAGG CCCAGCGGCT
 3120

GCGCCACACG CTCGGCCTCG ACCGCCCGCT CGCCGTCCAG TACACCAGCA TGCTGGGGCA
 3180

CGCCCTGCGC GGAGACCTCG GCACCTCGAT CCGCAGCGGG GCGCCCGTAC GCGGGCAACT
 3240

CGCCCAAGCG CTGCCCCACA CCCTCTCGGT GGCCCTCCCC GCGCTCGTCC TCAGCGTGCT
 3300

CGTCGCGCTC TGCTCGCCC TCCTCGGGGC CTGGCCCCGG CGACGAGCCC TGCGCAGGGC
 3360

GGCGACCGCC CTGCCCTCGC TCGGCACCGC GATGCCCAGC TTCTGGCTGG GGCTGCTCCT
 3420

CGCCAGTGG GTCTCCTTCC GCTGGGGACT GCTCCCCGCC ACGGGCGGCG GGCCTCGCC
 3480

CCGGGCCACC CTGCTGGCCG CGCTGACCCT CGCCCTGCCC ATCGGCTGCG TGCTCGCCCA
 3540

GGTGCTGGGG AGGGGCCTGC GCGCCGCGCT CGCCGAGCCG TACGCGGACG TGGCGCGCTC
 3600

GCGCGGCGCG GGGCGGGCCA GGCTGCTGCT CGCGCGCGCG CTGCGCAACG CCTCGGTGGC
 3660

CGCCCTGGCC CTGCTCGGCG TCGTGTGCGG GCAACTCCTC GCCGGGGCCG TCCTGGTGGA
 3720

GACCGTCTTC GCCAGGGGCG GCATCGGGCG CCTCGCCATG GACGCCGTCA CCTACCAGGA
 3780

CCTGCCCGTC GTGCAGGGCG TGGTGGTGCT CGCCGCCCTC GTCGCGGCGC TGGTCAACCT
 3840

CGTGGTGGAC CTCCTCCTCC CGCTCCTCGA GCCGCGTACC GCCTCGGAGG CCGCCGATGC
 3900

CGTCCCTGCC CACTGACTCC CCGGGCCCCG CCCCCTACCC GGCCCTCCC GCCCAGCCGG
 3960

TCACGGACCG CGCCGAGGGC CCGGAGCCCC GGGAGGCGGG AGCGCGCCGC GCGGGCCGCG
 4020

TGGCGCTCCG GCGCGTCGCC GCGCTGTGGC GGGCGCCCCG ACTGGCCCTG TCCCTGCTCG
 4080

TCCTCCTGCT CGTCCTCGGC TGGGCGCTGC TGCCCGGCCT GTTCACGGCG GCCGACCCTC
 4140

TCGGGGCCGA CGCCGCGCAC CGCCTGCTCG CCCCCGGCGC CGGACACCCG CTCGGCGCCG
 4200

ATCAGTCGG CCGCGACCTC TACGCGCGGG TCGTGCACGG CACCGCGCGC TCGCTGGGCA
 4260

CGGCGTTGCG CGCCGTCGCG CTCGGGGTGC TCGCCGGCGG TCGCTCGGG GCCGTGGCCG
 4320

GAGTGGCGGG GCGCGCGGTG GACGCCGTG TCATGCGCGT CGTGGACGTC CTCCTGGCCG
 4380

TCCCCGGGCT GCTGCTCTCG CTCGCGGTGG TCTCCGCGCT CGGCTTCGGC ACCGCGCAGG
 4440

TCGCCTGCGC GGTGCGCGTG GGCACCGTCG GCGGGATCGC CAGGGTGAGC CGTGCGCAGG
 4500

TCGGCGCGCT GCGCGGCGGC GAGTACGTCG AGGCCGCGCG CCTCGCCGGG GTCGCGGGGC
 4560

CGCTGATCCT GCTGCGCCAC ATCGTGCCCA ACGCGGCTCC GCCCGTGCTC GCGCTCGCCG
 4620

TCACCGAGTG CGGCACCGCG GTACTCGGCG TCGCGTCCCT CGGCTTCCTC GGCTTCGGCG
 4680

CGCCGCCGCC CGCCCCGAG TGGGGCGCCC TCATCTCCAC CGGCCGGGAC TACCTCGTCT
 4740

CCGCCTGGTG GCTCACCACC CTGCCGGGCC TCGTGCTCGT CGCCCTCGTC GTCGCCCTGC
 4800

ACCGCGTCGG CCGCGCCCTG GAACGAGAGG AACGCACCGG ATGACCCCGG CGGACAAGCC
 4860

GACCGACGAG CGCTCACCTG TCCTGGACCT CAGTGGTGTC AGCGTCGCCT ACGGGACGCG
 4920

CACCGTGCTG CACGGCATCG ATCTGCGCCT CGCGCCGGGG CAGGTGCTCG CGGTGCTCGG
 4980

CGCCTCGGGT TCCGGCAAGA GCACCCTCGC GCAGGCGGCT CTCGGCCTCC TGCCCCCGG
 5040

CGGCCGGGTG ACGGCAGGGC GCGTCACGGT CGCCGGGCAC GACATCACCG CGCTCGCGCC
 5100

GCACCGGCTG CGCGCGCTGC GCGGCACGGT CACCGGACTC GTGCCGCAGG ACCAGGCGGT
 5160

CTCCCTCGAC CCCCTCGTCC GCGTCGGCGC GCAGGTCACC GAGACGCTGC GCGCCCATCG
 5220

CCTGGAGGAC CGCCGCGAGG CCGCCCGGCG CGCCGTGCCC TTGCTCGGCG AGGCGGGGAT
 5280

CGAGGCCCCC GGGCCGCTCG CCCGCGCCTA CCCGCACGCG CTCTCCGGCG GCCAACGCCA
 5340

ACGCGTCCTC GTCGCCGGGG CCTTCGCGGC CCGCCCCCG CTGGTCGTCTG CCGACGAGCC
 5400

GACGAGCGCG CTCGACGCGA CCGTGC GGCG CCGGGTCATG GACCGCTTCG CCGCCCTCGT
 5460

CGCCGCGCAC GGCACGGCCG TCCTCCTCGT CACCCACGAC TTCCGGCTCG CGAGGGAACG
 5520

CGCCGACCAG GTCGCCGTCC TCGCGGACGG CCGTCTCGTG GAGAGCGGCC CCGCCGCACG
 5580

GGTCTCTGAC CGGCCCGCCC ACCCGTACAC CCGCCGCCCTC ACGGGGGCCG GCCGCCGCGT
 5640

GGCCGCACGG GGCACGGCGC CGCGCGCGTC GGGCACCCCC GTGGTGCGCG CCCGCGACCT
 5700

CGTCAAGGAG TACCGGCGGG ACGGGCGGCG GGTGCGCGCC GTGGACGGGG TCGGCTTCAC
 5760

GGTCCGCGAA GGCGAGTTCT TCGCCCTCGT CGGCGAGTCG GGCTCGGGCA AGTCCACGAC
 5820

CGCCCGGCTC GTGACGGGTC TGACCGCGCC CACGTCGGGC GCGGTGGAGC ACGCGCCCGC
 5880

TCCGGTGCGG CCCCAGCTCG TCCAGCAGAG CCCGTACGCG GCCTTCGACC CGCGCTGGAC
 5940

CGTGCGCCGC ATCGTGAGG AGCCGCTGCG GGC GCGGCAC GTGCCCCGGG CGCGGCGGCG
 6000

GGCGCGGCTG CGCGAAGTGC TCGCGCTCGT CGGCCTCGAC GAGGAGCTGC TCGCGCGGCG
 6060

GCCCCGGGAG CTGTCCGGCG GGCAGCGGCA GCGGGTGGCG CTCGCCCCGCG CGCTCGCCCC
 6120

GGAACCGCGC CTGCTCGTGT GCGGCGAACC CGTCGCGGCC CTCGACCCCG TCGCGCGCGA
 6180

GCGCGTCGTC CACCTCCTGG AGCGGCTGCG GCGGAGCTG GGCCTCACGT GCCTCTTCGT
 6240

CTCCCACGAG CTGGACGTCG TGCGGCGGTT GTGCGGTCGC GTCGCCGTCA TGCGGGGCGG
 6300

GCGGCTACTG GAGAGCGGAC CGGTGGGGGA GGTGCTGTCTG GCGCCCTCGC AGCCGTACAC
 6360

GCGGGCGCTG CTCGCCGCGG AGGCGGGCCC TTCCGACACG CCCGGAGCGG GGTGAGCTGC
 6420

GTCACGTCGG GAGGGGCTTG CGCGGCCCCC GCCAGCCCGC CTACGCGAAG GGAGGGGGCT
 6480

TACCTGGAGG TATGGGCTTA CCCGGGGAAA GTCTGCAGCC TAGGCTCGCC CCGTGAACGA
 6540

GATCACTGTC GAGATCTGGA CCGACGTCGT CTGCCCCGTG TGCTACATCG GCAAGCGGCG
 6600

CTTCGAGCGG GCGCTGGCCG CCTTCGACGC GAAGGAGGAC GTGCGCGTCC ACTGGCGCAG
 6660

CTTCGAACTC GACCCCGCCG CCCTGCGGGT CACCGACGAG ACCATCCCCG AGCGCATGCT
 6720

GCGCCGCCAG GGCATCCCGC CCGAGCAGGC CGCCGAACTC CTCGCCGGGG TGAGCGCCCA
 6780

GGCCGAGGCC GAAGGGCTGG AGTACCACCT CGACCGCGCC CGGCCCTGCA ACACCTTCGA
 6840

CGCGCACC GG CTCGCCACC ACGCGGGCAC GCGGGGCCTC GCGGAGACCT TCCAGGAACG
 6900

GCTGATGTGC GCCTACACCG CCGAGGGCGT CTCGGTGGGC GACCACCCCA CCCTGCTGGC
 6960

CCTCGCCGAG GAGGCGGGGC TCGACGCGGC CGCCGCCGCC GAGGTGCTGG CCGGTGACGC
 7020

GCACGCCGAG GACGTACGGG CCGACGAGGA CCGGGCCGCC CGGCTCGGCG TCGGAGGGGT
 7080

GCGCGCCTTC GTGATCGGCG GACGCTGGTC CGTCTCGGGC GCCCAGCCCG CCGAACTCCT
 7140

CACCGGTCTC CTGGAACGGG CCCGGACCGC AGCCGECGCC TGACGGTCAC CCGGACCGTC
 7200

CGCGGGGCGG ACGCGGCCGG GTACCGCGCC GCAGTCTCCG CACCCGCTT CACCCCTCCA
 7260

CGCCCGCACC GGTCCCGTCC GCGCCCGGGC CGTCCCGCCC GGGGGCCGGT CGCCCTTCC
 7320

GCCGTCCCAA GGAAAGCGTC GCCATGTCCT CCTCCCCGCC CGCCCCCGCC ACCCGGGCG
 7380

TTGCCCCGCA CTCGCCGCCC GCTCCGCGCC TCGGCCTCGT GCTCCTCGTG TGCTGCCTGG
 7440

CGCAGTTCCT CGTGACGCTC AGCGTCGCCA TCGTCAACGT GCGGCTCCCC GACATCCAGC
 7500

GCGGGCTCGG CTTAGCGCC GAGAGCCTCC AGTGGGTCGT CAACGCCTAC ACCGTCACCT
 7560

TCGCGGGCTT CCTGCTCCTC GGC GGACGGA TCGCCGACCT CTTGCGGGCGG CGCCGCATCT
 7620

TCCTCGCGGG TGTCGCGCTC TTCGCCCTCG CGAGCCTCGC CGGGGGCCTA AGCCAGAACG
 7680

CGGGGACCCT CGTCGCCGCG CGCGCCGTGC AGGGCCTCGC CGCCGCCGTC ATCGCGCCCA
 7740

CCACCCTCGC CGTGCTCGGC ACGAGCTTCA AGGACCCGCA CCAGCGCCAC CGCGCCTTCG
 7800

GCGCCTGGGG CGCGGTCTCC GGTGCGGGCG GCGCCTTCGG CGCCCTCGCG GGC GGCGCGC
 7860

TCACGGACGC GTTCTCCTGG CGCTGGGTGC TCTTCGTCAA CCTCCCGATC GCGTGCTCC
 7920

TGCTCGCCGG GATCGCCTGG GGCATCAGCG AGTTGCGGCA CGCGGGCGAG GACCGCCGCA
 7980

TCGACGTCGC GGGCGCCCTC ACCGTGACGC TCGGTCTCCT CGCGCTCGTC CTCGGCATCG
 8040

TGCAGAGCGG CCCGCACGGC TGGGGCTCCG CCGCCACCCT CGTCCCGCTC CTCGGCGGCC
 8100

TCGCCCTGCT CGGTGCCTTC GTTCTCGTCG AGGGGCGGTT CGCGCCGCAG CCGCTCATAC
 8160

CGCTCGGCAT CTTCCGCTCC CGCTCCGTCG TCGCGGCCAA CGTCGTGGCG ATGACCAGCG
 8220

GCGCCGCGCT CTTAGCATG TTCTACTTCC TCACCCTCTT CCTCAACCAG GTACGGGACT
 8280

ACAGCCCGCT GCGCACCGGC TTCGCCTATC TGCCGCTCGC GCTCGCCATC ATGGTCGCCG
 8340

CGCAGTTCTC CGCCGCGCTC GTGCGCGTCC TCGGCCCCCG TACGACCCTG CTCGTGTCCA
 8400

TGGCTCTCAC CGCCGCCGGA CTGCTCTGGC TCTCGCGGCT CACCGAGGAC TCCGGGTTCG
 8460

CGGGCGGACT CCTCGGCCCCG ACCCTGGTCG TCGGCATCGG ACAGGGCATC TCGATGTCCG
 8520

CCTCCGCGAT CGCGGGGGTC GCCGGGGTGC GCCCGCAGCA GGCCGGGCTC GCCTCCGGGC
 8580

TGCTCAACGC GACCCGGCAG CTCGGCGGCG CCCTCGGCCT CGCGGTCGTC GCCGCCGTG
 8640

CCACCTCGCG CGCCGACGGA CTCCTCGACG GGGTCGCGCG GCCACGGCG GAACTCGCCC
 8700

GGCACGCCCA GCGTCCGGG CACCCGCTCT CCATCGCCGT GGCCGCGGCG CTGTCCGCCG
 8760

TCGGCCTGCT CGCCTCGCTC GCCGCCCCGG GCCGCTCTCC GGCCCCCACC GGCACCCGCA
8820

CGGGCGGCGA CTCCGCCGCA CCGGCTCCCG CGGCGGCCCC CGCCGCCACC GGGACCACCG
8880

GCCCCGGCGA GATCTGAGCG CTCCCCCGG TGACCGCCCC CTCCGGCGGG GCGGTCACCG
8940

GGGACGGCGC GGCCTGCGTA TAAACATGCG CGGGCCTTTT GCATGTGTGG GAAAGATGCG
9000

GGGCGTGCAC ATAATCACCG GCTTGCGCGC CGCGACTGTG TGCTGCGTCA CGTTTCACGT
9060

GGTGTCTATA TTTCCCTGTC TATTCTTGGC ACCGCTCGGA GGGGGACCGG TGC GCGCTGC
9120

CGTTTATCCC GCCGGACGTC ATGATGTCCA CCGGGAGGGA AACGTGGAAA TCAAGTACAT
9180

GGGTCAGTTG ACCATGCGGT GGGAGGGGCG GGAAAAGCTC CCCTCGGCGC GCAAGCCGCG
9240

CACCGTGCTC GCCCTTCTGC TCCTCAACGA CAAGACACCC GTCACGACCA GCGCACTCAT
9300

CACGGAAGTC TGGGGCGAGA ACCCGCCGCG CAGCGCCCTG ACCACCCTCC AGACCTATAT
9360

TCTCCAGCTC CGGAAATGCC TCGCGGCGAT GAGCGGGCGA AGCCTCGCCT GCATATCCGA
9420

GAAAACCCTG GTGACGTGGC CCTGCGGTTA TCTGGCGCGG CTTCCGGCGG ACGCCACTTC
9480

GGATGTGGCT GAATTCCGTC GGT TTGCCAG GGAGGGGCGG GAAGCGGAAC GCAGGGGTCA
9540

TCTTGCCGAG GCCGTCCGCT CCTACCGCGC GCGCTGTCA CTGAGCCAGG GACCACTCCT
9600

CGCCGATATC GAGCACGGGC CGCTGCTGCG CGCCGAGGCG GTACGCATGG AGGAGTGCCG
9660

GCTTTCGCTC GTCGAACGCT CCATCGAGGG CGACCTGCTC CTCGGCCGGC ACCGGGAGGT
9720

CGTCAGCGAA CTGTCCGCCC TCGTCGCCCA GTACCCCTAC CACGAACAGC TCACGGGGCA
9780

GCTCATGGTC GCGCTCGTGC GCTGTGGCCG CCGGCAGGAC GCGCTCGCCG TGCACCAACG
9840

GCTCCGGGCC CGCATGGTGG AGGACCTCGG CCTCGAACCG AGCAGCCACC TGCGCGCCCT
9900

GCAGTCCGCC GTCCTCAGCG GCGAGCCGCT GCCGGGACCG CCCGGAACCG GTGGCGAGAT
9960

CCCCACCCCC TACGCGGGCG CTTTCGCCAC CGCGGCGCGC TCGCACGACT GACCCCGCCC
 10020

GGGGCCCCGT CCACCCGTCTG CCGTCGCCGG GTGGAAGGGC TGGACGGAGA ACAGGCCCTC
 10080

TCCTTCCCGC CGCCGGGGCC CGCGCTCAGC GCAGCGCGCG AAACCCCCGC TCCCAGTGCA
 10140

CGAGCGGTTC GCCCTCACCC TGCCCCACCC CGCGCACCTC GCCGACGAGG ATCACGTGGT
 10200

CCCCggccag GTGCCGGGCG TGTACCGCGC ACTCCAGCGT GACGAGCGCG TCCGCCGCGA
 10260

CCAGCAGCCC CGAGCCCTCC AGCGCCACGA GCCCCCGCC CCCGAACTTG TCCGCCCCGC
 10320

GCGTCGCGAA CCGCGAGGCC AGCTCCCGCT GTCCGCTCCG CAGCACGCTC ACCGCGAACC
 10380

GCGCGCACTC CTGGAACGCC CCGTAGGACT GCGCGCCCGT GCCCTGGCAC ACCGACACCA
 10440

GCGGGGGAGC GAGCGACACC GAGCAGAAGG AACTCGCCGT GAAACCGCGG GGGGTGCCGT
 10500

CCTCCTCGCG CGTCGTCACG AGGACCACCC CCGCGGGGAA GCACGACATC GCGGCGCGGA
 10560

AGCCCTCGGC ATCGACCGGC CCGACCTGCC CCGTCCGCGC GGCCGTCCGT GTCTCGTTCA
 10620

CGGGAGCTCC ACCACCTGGG CCGCGTAGTT GAGCCCCGCG CCGAACCCCA CGAGGAGCGC
 10680

GCGGTCCCCG CTGCCCACCT CGCCGGAGTC GAGGAGCGCC TCCAGCGCGA GCGGGACCGA
 10740

GGCCGCCGAG GTGTTCCCGG CGCGCACCAC GTCCCGCGCC ACGGCCGTGC GCTCCGGCAG
 10800

GCCGAGCCGT TCGGCCATCA GCTCGATCAT GCGCAGGTTC GCCTGGTGCG GGACGAAGGC
 10860

GTCCAGCGCC TCCGGCTCGA TCCCGCTCTC CCGCAGCAGT TCCGCCGCC GCGGGGTACAC
 10920

CTCGTCCATC GCCCAGCGGA ACACCCGCCG TCCGTCCATC CGCATCCACG GCCTGCCGAG
 10980

CGAGGGATCG GCGGCGAACG TGTCCCAGCC CCTGTCCATC CGCAGCGCTC CGGCGTACCG
 11040

GCCCGCCGCA CCACGCGCCG GCGGCGAGAT CCCCggccgc GCCGAAGGCC CCACCACAGC
 11100

GGCTCCCCGC CCGTCCGCGA AGAGCACGGA GACGGTGCGG TCCGCACGCT CCACGAGATC
 11160

CGTCATCCGC TCCGTGCCCA CCACCAGGAC GTGCCGGGCG CTCCCGGCGC GCACCGCGTC
 11220

CGAAGCGAGC GCCAGCGCGT GGCAGAAACC GGCGCAGGCG CCGGACACGT CCAGCCCCGC
11280

CGCCGCCCCG GCGCCGAGGT CCTCGGCGAG GAGCACGGCG AGCGGGGGCG TCTGCTCCAG
11340

CCGCGACATG CTCGCCACCA GCACCAGATC GATCCCGTCC GGCAGCACCC CCGCCCGCGC
11400

GAGCGCCTTC TCGGCCGCCG CCCGCCCCAT GAACCGCAAG GTCTCGTCCG GCGCGGCGAA
11460

GCCGCGCGCA CGGATACCGC TCGGGGTCTC TATCCACTCC GGGGTGACG CGACCGTGGC
11520

ACACAGTTCG GCGTTGTCCA CACTGCGCCG GGGCCGGTAA CCCCCACCC CCAAGAGCCT
11580

GCTGTGCACG ACAGGCCGCG TACCGAGCAT CTCCGCCTCG CCGGTCACTG TCCCTCCCTC
11640

CCGCCAGGGG CCCGGTGCCT CCGGCCGCTG ACAGCCTCCG GGCCGCCTCT TGAGCCCGGA
11700

TCGAGCCGAG ATCGAGACCC CGGCCGGATG CTGGGGCCGT CACTCCCCGG CGCCGGCCGG
11760

GGGCCGAGGG GACCGGGCTC CGGGTCCCCA GGAAGGAACA CACGCGGCAT GAGCGCCTTC
11820

ACTCTCACCG AGTTCAAGAA GCTGGTCGAG CAGAGCTACG ACGCCGAATC GGCCGAAGCC
11880

CTCCACGGGC AGGCCCTCGA CACGAGCTTC ACCGATCTGG GTTACGACTC GCTGACGGTC
11940

TACGAGATCG TCACCCGCAT CCAGGACGAG CACGGAGTCA CCGTGCCCGA CGAGGAGCTG
12000

GACCTCCTCG ACACCCCGCG CGCCCTCATC GCGTACGTCG ACGCGCGCGC GGGTTCCCGC
12060

ACCTGACCCG ACGGGAGGGT GCGGGGAGGG CGGGGCCGGA GCGGGCGGGG CGCGGCGGCG
12120

GTACGACGCC GCCGCGCCCC GGGCGGCAGG CGCTCAGCCC CCGTCCCGTA CGACCTGACC
12180

CGCCTGTTGG AGCGCGGCGC GAAAGGCCCG CACGTCGTCC GCGGGGTGCG CGGGACGCGC
12240

GGGCAGGAGC CGAACGACAC GGGCCGCGCC CGAGCGGACG GCCGGATGAC GGCGCGCCGG
12300

GGCGCCGAGA CTGCCGTCCG GCGAGGCGTC GAGCAGCACC ACGCCACGCC CGCCGCCCGT
12360

GGTCGTCGCC GTCGCCGTCG CCGTCGCGAG CAGGTTCCCC AGCGCCGGCC AGTACAGGAC
12420

GGGGGCGGCG AGCTGACCGC ACCAGAACGC GGC GCGCACC GCCTCGCCCC GCGTGAGCGC
12480

GCGACCCGTC CGGCCCCACC ACAGGGGCAC GGC GGGTGGC CGCCACCGCG CGACCCGTTT
12540

CGCGCTCGCC CTGGTCAGAC GCCGCGCCGC CCCGCGGAGG GCGGGGGAGT GGAAGGGTAC
12600

GTCCGCCTGG AGGCGGCGCG CCGTCAGCCC GTCGTCGGCC AGGCGCCGCG CCACCGCGCG
12660

CAGCGGCGCC TCCGGCCCCG ACAGCACCGT CTGACGCGGC CCGTTCACCG CGCCCACCAC
12720

CACGTCGGCC CCCACGTACG GGC GCGAGTC GTCCACCGGC GCGGCCACGG CGAGCATGCC
12780

CCCGGACGGC GCCGCGCGCA GCACCGCGTC CCGTTCGGCG AGCAGCCGGA GCGCCGCCCC
12840

CGGCGCGCAC ACCCCGGCGA GCGCGGCGGC GGC GAGTTCG CCGACGCTGT GCCCGAGCAA
12900

CAGGTGCGGC GCACCCTGCG CCCCCGCGCC GAGCGCCCGC CCGAGCCCGT ACCCCACGGC
12960

GAAGAGCAAC GGCTGCGCGA CAGCGGCCCC TTCCAGCGCC TCCTCGCCGC CCGCCCCGTC
13020

CAGCCACCGC GCGCGCAACT CCTCCCCCA GGTGCCCCAC ACGGCGAAGC ACTCCTCCAC
13080

CGCGGTCTGT AACTCCGGCA CCACGCCGTG GAGTCCGGCC GCCATCCGGG CCCGCTGCGC
13140

GCCCTGACCG GGCAGCAGCA GGGCCACCAC CGGGGCGCCC TCCTCCTCCG TACCGCCCTC
13200

GCTCCGCACT CTTCCGCCTC CTCGTCCCGC CGACTCCTGC CCGTACCGAG CCTCGCCCTT
13260

GCCGCTCGAC GGCCGGTGGA CTCCCCTCC AGCGCGGCCC GCCCGGGGCG CCACCGAGCC
13320

TCAAGCCCGC TTCGAGGGAC GTCTTCCAAG GTCGGGTGGA GGCCGTGAAC GGCACCGCGC
13380

CGTCCAGGAC GCCCAAGGGG GTCGTGCAGT GAACCGACAA GTCGCCGTCA CCGGCATCGG
13440

CGTGGTGGCT CCGGGAGGGA TCGGGCGCAA GCCGTACTGG GAGCAGCTCA CCTCCGGACG
13500

CACCGCCACG CGCGCCATCT CCTTCTTCGA CGCCTCGCCC TTCCGCTCGC GGATCGCCGC
13560

CGAGGTCGAC TTCGACCCCG CGGCGGCGGG CCTGAGCCCG CGCGAGGTCC GCCGCATGGA
13620

CCGGGCGCGC CAGTTCGCCG TCGTGAGCGC CAGGGAGAGC CTCGCCGACA GCGGACTCGA
13680

CGTCGCCGAC CTCGACCCCC ACCGGATCGG GGTGAGCATC GGCAGCGCCG TGGGCGGGAC
 13740

CACCTCGCTG GAACGCGAGT ACCTCGCCCT CAGCGACAGC GGGCGACAGT GGGAACTCGA
 13800

CCTCTCCTAC CTCTCGCCGC ACCTCTACGA CGCCTTCACC CCCAGCTCGC TCGCCCGCGA
 13860

GGTGGCCGGG GTGATCGGCG CGGAGGGGCC CGCGGCAGTC GTCTCCACCG GCTGCACCTC
 13920

CGGCATCGAC TCGCTCGGCC ACGCCCGCGA CCTCATCGCC GAGGGCAGCG CCGACGTCGT
 13980

GCTCGCGGGC GGCACCGACA CCCCCATCTC GCCGATCGCC GTCGCCTGCT TCGACGCCAT
 14040

CAAGGCCACC TCGCCAGCA ACGACGACCC GGCGCACGCC TCGCGCCCTT TCGACAGGGA
 14100

ACGCAACGGC TTCGTCCTCG CGGAGGGCGC CGCCGTCTTC GTGCTGGAGG AGCTGGGGCA
 14160

CGCGCGGGCC CGCGACGCC ACGTCTACGC TCTCGTGTCC GGCTACGCCA CGCGCTGCAA
 14220

CGCGTACCAC ATGACCGGGC TCACCCCGCA CGGCAGGGAG ATGGCCGAGG CGATCCGGCA
 14280

CGCGCTCGCC GAGAGCGGGA CCGACCCCGC CGCCGTGGAC TACGTCAACG CCCACGGCTC
 14340

CGGCACCAAG CAGAACGACC GGCACGAGAC CGCCGCCTTC AAGGCCACGC TCGGCGAGCG
 14400

GGCCAGGAGC GTGCCGGTCA GCTCCATCAA GTCGATGATC GGCCACTCGC TCGGTGCCAT
 14460

CGGTTCGCTG GAGATCGCGG CCTGGGCGCT CGCCATGGAG TACGGCGTGG TGCCGCCAC
 14520

CGCCAACCTC GACACGCCCC ACCCCGAGTG CGACCTCGAC TACGTGCCCC ACGAGGCGCG
 14580

CGAGACGCGG GTGGACAGGG TCCTGAGCGT CGGCAGCGGC TTCGGCGGTT TCCAGAGCGC
 14640

CATGGTCCTC ACCCGGGACA CGGGCGCCCC CCTCCCGACG GCCTGACACC CGCTCACCGC
 14700

ACGCCCCGCA CCGGCCCGGA CGGCCGGCCA GGTCCCGCAC CGCTCGCACA CGGAAGGCAC
 14760

CCGCCCATGA CCACGCCCC CTCCCGCACC GCCCAGGGCG CCCCTCCCGG GGCCGCCCTG
 14820

CCCCCGTCT TCACCGGCAT CGGCGTCGCC GCCCCAACG GGCTCGGCAC CGAGGAGTGG
 14880

TGGGCGGCCA CCCTGCGCGG GGAGCACGGT CTGCGGCCCCG TCACGGAGTA CGACGCGAGC
 14940

GGCCACCCCG GTGGACTCGT CGGCCGCGTC CCGGACTTCG ACGCCGCCCCG CCACCTGCCC
 15000

GGCCGCCTCC TGCCGCAGAC CGACCGGGTC ACCCGGCTCG CGCTCGTCGC CGCCGACGAG
 15060

GCACTGAAGG ACGCCGCCGT CGACCCGGCC CGGCTGCCCCG AGTACGGGGC GAGCGCCGTC
 15120

ACCTCCAACG CCACCGGCGG CTTGAGTTC ACCCACC GCG AGATCCGCAA ACTCTGGACC
 15180

GAGGGCCCCG CCCGCGTCAG CGTCTACGAG TCCTTCGCCT GGTTCACGC CGTCAACACC
 15240

GGCCAGATCT CCATCCGGCA CGGCATGCGC GGCCCCGGCG CCGTCGTCGT GGCCGACCAG
 15300

GCGGGCGGCC TCGACGCCCT CGGCCAGGCC CGCCGCGTAC TGCGCAAGGG CGGGGTGCTG
 15360

GCGGTGAGCG GCGGCGTGGA GTCCGCGCTC GACCCCTGGG GCCTGGCCGC CCACGCCTCG
 15420

TCGGGCACCC TCAGCCGCTC CGGCGACCCG GCCACCGCCT ACCTCCCCTT CGACCGGCGT
 15480

GCCCTCGGCA CCGTCGTCGG CGAGGGCGGC GCCCTCCTCA CCCTGGAGAC ACCCCGGCAC
 15540

GCCGAGGAGC GCGACGCACC CCGGATCTAC GGCGAACTCG CCGGGTACGC CGCCACGTTT
 15600

GACCCGCCCCG CGGGCTCCGG ACGGCCCCC GGCTCGAAC GCGCCGCGCG CCTCGCCCTC
 15660

GCCGACGCGG GCCTGGCACC CGGGGACGTC GACGTCGTCT TCGCGGACGC GGCGGGGCTC
 15720

CCCGCCGCCG ACGCCGCCGA GGCCGCCGCC CTGCGCGCGC TCTTCGGCCC CGGCGGCGTT
 15780

CCGGTGAGCG TGCCGAAGAC CCAGACCGGG CGGCTCGCCT CGGGCGGCCC GGCCCTCGAC
 15840

GTCGCGGCCG CCCTCCTCGC CCTGCGCGAC GGCTCGTG CCGCGGCCGT CCACCTCGAC
 15900

GAGGTCGATC CCGCGTACGG CCTCGACCTC GTCCGCGACA CCGCGCGCGC CCTCCCGCTG
 15960

CGCACGGCGC TCGTCCTCGC GCGCGGCCAC GCGGCTTCA ACGCGGCCGT CGTCGTCCGC
 16020

GGGCGGCGGC GGCCCCGTAC CGCCTGAGGG CCTGCCCCGG GTGCGCCGCC GCGGCCGCAC
 16080

CCGCCGTCCA CCGCCCCGCA CACCCACAA GGAGTTCCCA TGAGCGCACT GACCGTCGAC
 16140

GACCTCAAGA AACTGCTCGC CGAGACCGCC GGGGAGGACG ACAGCGTCGA CCTCGCCGGA
 16200

GAACTCGACA CGCCCTTCGT GGACCTCGGC TACGACTCCC TCGCGCTGCT GGAGACGGCC
 16260

GCCGTGCTCC AGCAGCGGTA CGGCATCGCG CTGACTGACG AGACGGTCGG CCGGCTGGGC
 16320

ACCCCGCGCG AACTGCTCGA CGAGGTCAAC ACCACCCCGG CCACCGCCTG AGCGGGCGGC
 16380

GCACGCGAAC GAAGGGTCCG GCCCCACCCC CCGCGCGGGG GGTGGGGCCG GACCCTTCGT
 16440

TCGCGGCGGA CCTCACAGCA AGGTGCGGTC CTGGGTCACC TGCCGCGACA GCGTGCGCAG
 16500

TTCCCCGCCC TCCCGCACCA GGACGTCCCG TACGACGCAA CTGGGACCGA TCACCGGGAG
 16560

CGCCGCGTCC GGCCGGGTCA CCACGACCAG CGCGTAGAAC TCCGTCTCGA CGGCGCCGTC
 16620

CTCACCCGGG CGTACGTCCA GCATCGAGAA CCAGTGCCGC CGCTGGACGG GGTCGTCGGC
 16680

GAACCGGCGG GCGTTGAAGC CGCGCACCTC CTCCGCGATC GCCGCCCTGC CCCGGGCCGC
 16740

CGGGGTGTCG GGGGAGTGGG CGAAGACGCC CTCCGCCGCG AAGGTGGCCG CGAACTCCTC
 16800

GGCCCGTACC GCGTCCAGGT GGTGTGCCTG GCGCGCGTAG AACGCCTGTA CCTCGGTGTA
 16860

GAGCGCCGCG GAACGCGCCG CGTCGCCCCG CAGGGACCCG GCCCTCACGT CGTCCGGCGG
 16920

GGTGGTGGTC GTCACGGGAG CCTCCCAGAA GGTGTCTGAC CTGTCCCGGC AAGCCTCCGC
 16980

GCCGTGGCTC GACGGCCCCCT TGAGCCCCC TCAAGGGAGC GCCGCGAGGC TGGCACCAGC
 17040

ACCGCGGCGT CACCACCGAC GCCCCCGCGC CGAGGAGGAA GGCCATCATG ACCACAGCAG
 17100

CTCCGCACAC CCGCCCCGGG GAGGCAGGCA CCACCCGGGG ACCCGCTCTC GTGACCGGCG
 17160

CGACCCGGGG CATCGGCCTC GCCGTGCGCG AGGCGCTCGT CGCGCGCGGC TATCCCGTGG
 17220

TGGTCTGCGC CCGCGACGCC GAGGCGGTCTG CGCGCACCGT CAAGGAGCTG GCAGCGGGCG
 17280

GCGCCGCGT CGAGGGCGTC GTCGCCGACG TCACCGACGC CGCCTCCGTG CACGAACTCG
 17340

TCGCCACCAC CGTCGCCCCG TTCGGCCCCG TCGAGGTCCT CGTCAACAAC GCGGGCCGGT
17400

CCGGCGGGCGG AGTGACCGCC GAACTCAGCG AGTCCCTGTG GGACGACGTC ATCGCCACCA
17460

ACCTCAAGAG CGTCTTCCTG GTCACCCGGG AGGTGCTCAC CACGGGCGGG ATGACCGGGC
17520

GCGGCCGCGG CGTCGTCAAC ATCGCCTCCA CGGGCGGCAA GCAGGGCGTC GTCTTCGGCG
17580

CCCCCTACTC GCGGTCCAAG CACGGCGTCG TCGGCTTCAC CAAAGCCCTC GGCCTCGAAC
17640

TGGCCCGGAG CGGCATCACC GTCAACGCCG TCTGCCCCGG CTACGTCGAG ACGCCGATGG
17700

CCGCCGGAGT GCGCCGCCAC TACGCCGACC TGTGGGACGT CACCGAGGAG GACGTGCTGG
17760

CCCGCTTCGA GGC GAAGATC CCGCTCGGCC GGTACACCCG CCCCAGCAGG GTCGCCGCC
17820

TCGTGCGACTA CCTGGTCACC GACGCCGCCG CGGCCGTCAC CGCCCAGGCC CTCAACGTGT
17880

GCGGCGGACT GGGGAAC TAC TATGACCAC CACCACGACC GGGCACCAGC GCCCCGGCTC
17940

CGCCGAACAC TCCGCGCGCC TCGCCGCCCC GCCCGCCTCC GCCTACGAAC TCGTCGCCGA
18000

CGTGACGCGC TGGCCCCTCC TCTTACCCCC GTGCCTGCAC GCCGAGGTGC TGGAGAGCGG
18060

TCCCGGCACC GAACGCGTGC GGCTGTGGGC CCTCACCGGC GAGCAGGTGC GCGGCTGGAC
18120

CTCGCGGCGC ACCCTCGACA GCGAGGGCCT GCGCGTCGGC TTCCGCCAGG AGGACAGCGC
18180

CCCGCCGCTC GCCGCGATGG GCGGGGAGTG GCGCTTCACC GAGGAGGGCG AGGACACGCG
18240

CGCCGTACTC GCGCACGACT GGACGCTCAC CGAGCCGGGC GCCGCACCGC ACCGCTGGGT
18300

CACCGAGACA CTCGACCGCA ACAGCACCGC CGAGATCGGG GCCGTGACCG CCTGGGCCGC
18360

GCGCACCCAC GCGGCGGGCG GCGCCGACGC GCTGCTCTTC TCCTTCACCG ACAGCCTCGA
18420

CATCGCCGCC CCCGCGCCCG ACGTGTACGC CTTCTCTGAC GCCGCCGACC AGTGCCCCGC
18480

GCGGCTCCCG CACGTCAGCC GGGTCGCCTT CTCGACCACC CCGGCGACGC CGCTCACGGC
18540

GGGCGCGGAG GTGCAGCACC TGGAGATGGA GACGCGCGCG GACGACGGCA CGCGCCACCT
18600

CACCCGCTCG ATCCGCCTCG GCTTCGCGGG CCGTCTCCTC GTCTACAAGC AGACCACGTT
 18660

GCCCCCCCCG CTCCTCGGCC ACGCCGGCTC CTGGGCGCTG GAGCCGCTGC CCGGCGGCGG
 18720

CACCCGGGTC ACCGCCCCGC ACCGGGTGGC ACTCGACCCG GACGCCGTGA CCGAACGCTT
 18780

CGGCGCGGGG ACCACGCTCG CGGCGGCGCG GGACACGGTA CGGGCGCTGC TCGGCGGCAA
 18840

CAGCAGGCGC ACCCTGGAGG CGGCCCCGCG CCACACCGAG TCGGCGGGGG AGCGATGACC
 18900

ACGACCGGCC CCGCGACTCC CCTGGCCCCC GGGCCCGCAT CCGCCGAGAC CGTCGCGCTC
 18960

GCCGACGGCG TGCACTCCTG GCTCCAGCCC GACGGCGGCT GGTGCGTCAG CAACGCGGGC
 19020

ATCTTGCTCG CGCCCGACCG CGTCGCGCTC GTCGACACCG CCGCCACCGA GGCGCGCTCC
 19080

CGCGCGCTCG GCGCCGCGGT CGCCGGACTC AGCCCGCATC CCGTGCGACT GCTCGTCAAC
 19140

ACCCACTTCC ACGGGGACCA CAGCTTCGGC AACGGCATCC TCGGCAAGGA CGCCGTCATC
 19200

GTCGCCCACG AGCGCACCCG CACCGAGATG GCCGAGGCCG GACTCGGGCT CACCGGACTG
 19260

TGGCCGGGCG TGGACTGGGG GCACGTGGAC CCGGTCCTGC CCCAGCTCAC GTACCGCAGG
 19320

CGCCTCACGC TCCACCACGG GGACCTCCGC GTCGAGCTGA TCCACCCCGG CCCGGCGCAC
 19380

ACCACCAACG ACACCCTGGT GTGGCTCCCC GAGCAGCGCG TCCTGTTTCG GGGGGACGTA
 19440

CTCCTGCCGG GGGCGACGCC CTTCGTGCTG ATGGGCTCGG TGACCGGCTC GCTGCACACC
 19500

CTGCGTCTGC TGCGCCGCCT CGGCCCCCGC GTCGTCTGTC GCGGGCACGG CCCCTCGCG
 19560

GGGCCCGAGG TGATCGAGGA GACGGAGCGC TACCTGCTGC GGCTGCGGCG CATCGCCACC
 19620

GAGGGCCACG CCGCGGGGCT CACCCCGCTG GAGGCGGCAC GGCGCCACGG CCCCAGGGCC
 19680

TTCGCCCACT GGAGCGAGCC GGAACGACTC GCCGCCAATC TGCACCGGGC GTACGCCGAA
 19740

CTCGGCCCCG CACCACTGGG CACGCCGCTC GACGTCTCTG CCTGCTTCGG CGACCTCATC
 19800

GCCTACAACG ACGGCGAGCT GCCGGTCAGC CACGCCTGAA CAGCCCGGCC ACCGGCCGCT
19860

CACGGACCGC GCCGCGCCCT GCGCCGGTCC GCGCCGCCCC TGCCCGCGCG AACGACCGGC
19920

GCGGGCCGCC CCCGAGGGGA ACGGGCCGCG CCGGTCATCC GTGCGGACAG GGCGAGCAGG
19980

CCGCGCCGCT CACCCGCCGA CCACGGTGGG CGGTCCGGGC GCGAAGCTCG CCGCGTGGTC
20040

GCGCACCCAA GTGGCGAAGG ACCGCGCGGG CCGACCGGTC ACGTCGGTGA CACCCGGGCT
20100

CACGGCCGGT GGGTGCTCCG TGAAGAACCG CTCGACGCCC GCGAGCGCCC GCGCGTACGG
20160

CTCGGGGATG TGCCGGGCCG TCAGCTGCAC CCACTCCTCC TCCGGCACCG GCTCCACGCG
20220

CAGCGGGCGG CCCAGTTCCG CGGCGAGCAC GGCGGTCCGC CCGGGCAGCG TCAGCACCTC
20280

GGGACCCGAG AGCGGCACCG CCTTTCCGGC CCACCGCTCC ACGGGCGCCG TGAGAACCGC
20340

GGCGGCCACG TCGGCGATGT CCCGCGGGTC GATCAGCGGC ACCGCGAGCC CGGCGTGCGG
20400

GACCCGCACC ACACCGTCCC CGTGGAGCGC GGGCGCCCAC AGCAGCGCGA GCGAGGCGAA
20460

GGCGCCCGGA CGCAGGAACA CCCAGGGCAC CCCGCTCCCG CGCACGTGCC GCTCGCCGCG
20520

GACGTGCTTC TGGAACATCG GGTGTGCGTA CTCCGGGTGA GCCGCCGCCG TCGCGGAGAG
20580

CACCACCACC TCGCGCAGCC CCCGCGCCTT CGCCGCCGCC CGCGCGAAGG GCGCGCCGGC
20640

CTCGTCCTGG AGCAGCAGGA ACGCCCGCTC CACGCCGTGC AGCGCCGCCG CGAGCGAAGC
20700

CGTGTCCGCC AGGTCCCCGA CGACCGTCTC GGCGCCGACC GCGGCGAGCG CCCGCCCCGCG
20760

CGCGGCCGAG CGCACCAGGG CGCGCACCGG CACCCCGGAG TCGAGCAGCC GGGACACGAG
20820

GCGCCCCCGG ACCGCGCCGC TCGCCCCGGC CACCAGCACT GTTCCCGGCA GTGGTCCCCG
20880

CATCGTTCCC GTCACCCTTC CCGTCACCGG TTCGCTTCCT GTACCGCCCA GGCGCCGTAG
20940

GCGGCGCCCG CGGCCTCGGG GGAGAGCGCG ACGTGCGAGG CGCCGGTCTG CACGTCCCGC
21000

CAGAACCGCT GCACGGGGTC GTGCGGGCTC TGCCCGGACG TCCCGCTCGC ACGGACGAGC
21060

CGCTGCACCG CCCCACGGC CAACTCGGCG GCGAGGGCCG TGTCCTCGCTT GCCGCGCACC
21120

GCGGCGGCGG GCTCGGGAAG GGAAACGGTG CCGTCGGCCG CCGCCGCCGC GCGCAGCAGC
21180

AGCAGCTCGG CCGCGTCGGC CTCCCCGGCC GCCCGGCTCA GCTCCCCGAG TGCTCCGGGC
21240

GAAGGCGCCC GCCCCTGCGC GGCGCGCTCG GTCCACGCCC GCAGGGCGCC CCGCGCCGCG
21300

CCGACGAGCG GCGTCACGAC GGGCAGTGCC CCGACCCGCG CCATCCGCAC CGTGGGCACC
21360

GCGCCCGCGA CGCGGGGGCG CTGCCCCGCG AGGACACGGG TGCGCGGGAA GGAACGGTGC
21420

GCGGGGACGA GGACGTCGTC GAGCACGAGG GTGTCGCTGC CCGTCGCGCG CATCCCCGTC
21480

GTGTTCCAGG TCCGTTGAC GCGGGGTGCC GCGCGCGGCC CCGCGAAGAA CCGGACCTCG
21540

GGGCCCTCCT CGCCCTCGCC GCCCGGCACG GTCGCGCAGG CGAGCACCCA GCGGCGGTGC
21600

CGCACGCCGC TGATGTAGGG CCAGGCCCCG TCGAGCCGCC AGCCGCCGGG GACGGGCGTC
21660

ACGGTCCCCG AGGGGACGAG GGCGCCCGCG AGCAGCGCGT CCGGGCCCTC GGCCACACC
21720

TCGGCTGCC CCTCCTCGGG CAGGTACGCG GCGTACCGCC CGGCGTAGGC GGAGAGCGAG
21780

GCCGCCACG CGGCCGAGGT GCAGCCCGTC CCGAGGCGCA GGACGGCCTC GCGAGTTCC
21840

CCGAAGCCGC CTTGGGCGCC GCCGTGGGCC GCGGGCACGA AGTGGCGGGC GAAGCCCGCC
21900

GCCACGATGC CCGCGGACGA CTTCCCCGGC GAGACCCCGC TCCTCCTCAC GCGGCGCGC
21960

GTGCCCAGCC GCGAGCGCGG CGAGCGCCCC GGCCCGTTCC GCGAGCGGTC CGGGGGGCGT
22020

CCCGGCCCTT CGTCCGGCGC CCCGGTCCGC TCGGCGTGTG CCGGTGCGGC GAGGGTGGTC
22080

GATTGCTCCG TCATCGGTAC GTCACCTCCA CGCCGAACAT CTCCCGGCTT GCTGGAGGCG
22140

CGCTCGACTC GCGCTGGACG AGGCGGGGCG GCGAGCCCTC AGGCGAGGGA GTCCCCGGCC
22200

GGTGCGCCGG GCGTGGGCAC CGGGGCGTAG ACCTGCGGCG CGGTGTGGGC CTCGGCGCGC
22260

AGCCGCGCGA AGGCGGGCCG GATGCCGGGG GCGCTCGCCG CCGCCCGGTG GGCGGCGGCG
 22320

TCGCTCCAGT CGGCGAGCAT GGCCACCCG TCGGGGTGGT CGGGCGAGCG GTGCAGGCGC
 22380

AGGGAGAGGA AGCCGGGCTG CGCGCGCATG TGGGCGCAGA TCTCTTCGAG GACCGCGAGG
 22440

AAACGGCTCT CGTCGCCGTG CAGGAGCATC TGGTTGATCA TGGTCACCAT GTGTCCACCC
 22500

GACACCGTCC CGCTTGAGTC CCGCTTGAAG AGGCACCGCC GACACGCGGG CGGGTGCGGG
 22560

GAGTGAGGGG TACGGAGGCG TCCGGTGACG CGTACGGGGC GTCCGGTGAC GCTTTCGGAG
 22620

GCGGCCCGTG ACGCGCGGGG GCCGCCGGTG GGGGTTTCTC ACCGGCGGCC CCGGGCCGTG
 22680

CGTGCTCAGC CGAGCGGTGC GGGGAAGGTC GGGTACTCGA CGCCCGAGAC GTGCTGGACG
 22740

ACCCGTACGA CCTGGCACGA GTAGCCGAAC TCGTTGTCGT ACCAGAGGTA GAGGATCGCG
 22800

TTGTCGCCGT CGACCTTGGT CGCCCCGGCG TCCACGATCG AGGCGTGGCG CGAGCCGACG
 22860

AAGTCGCTGG AGACGGCGTC GGCGGAGGTG GTGAAGTCGA TCTGGCGGCG CAGCGGCGAG
 22920

GTGAGGGAGA CCTCGCGCAG GTGGGTGAGG ACCTCCTCGC GGCTCGTCCC GCGGCTCAGC
 22980

CGGAGGTTCA GGATCGCGAT CGACACATCG GGGACGGGAA CGCGGATCGA GCTGCCGCTG
 23040

ATCTTCGCCC CGAGGTCCGG CAGAGCCTTC GCGACGGCCG AGGCGGCACC CGTCTCGGTG
 23100

AGGACCATGT TGAGCGGCGC GGACCGCCCT CGCCGGTCGG ACTTGTTGTA ATTGTCCAGC
 23160

AGGTTCTGGT CGTTGGTGAA GGAGTGCACC GTCTCGACGT GGCCGCTCTC GACGCCGAAC
 23220

TCGTCGGCCA TCGCCTTGAG CGGCGGCACG ATGGCGTTTC TCGTGCAGGA GGCGCAGGAG
 23280

ATGATCCGCT CGTCCGGCTT GATCGTCTCG TGGTTCACGC CGTGCACGAT GTTCGGCACG
 23340

TCGCCCTTGC CGGGCGCGGT GAGGACGACC TTCGCGACAC CCGGGCGCAG ATGCTGCGAG
 23400

AGGCCCAGCG GGTGCGGCCA GCGGCCCGTG TTGTCGACGA GGATCGCGTC GTTGATGCCG
 23460

TAGGCGGTGT AGTCCACGGT CGCCGGGTCT TCGGAGTAGA TGAACCGGAT CGCGTTGCCG
 23520

TTCGCGATGA GGGTGTCTGTT CTCCTCGTCC ACGATGATCG TGCCCTGGAA CTGGCCGTGC
 23580

ACGGAGTCGC GCGCAGCAG TGAGGCGCGC TTGACGAGGT CCTGACCGGC GGTCTGCGG
 23640

ACCACGACGG CGCGCAGACG CAGGCCGTG CCCGAGCCGG CCTTCTCGAT GAGCAGCCGG
 23700

GCGAGGAGCC GCCCGATCCG CCCGAAGCCG TAGAGGACGA CGTCGCGCGG GCGGGCGCGC
 23760

TCGATCTTCC GCTCGCCGGT CGCGCCCTCG ACGGGCCGCG CGGTGAACTC GGCGACGCC
 23820

AGGCCCCGGT CGTCGCTCTT GTAGGTCTCG GCGAGCATGC CGATGTCGAT CTGGGAGGGG
 23880

CCGAGGTCGA GCGTCGTGAG CGCCTGGAGG AACGGCATGG TCTCGGTGAC CGACAGCTCC
 23940

TCGCCCGCGA TCTGCCGGGC GAATCGGTGG GTCTTGAGAA TCCCGACCAC CGACTTGTTT
 24000

ACCAGCGAGC GGCTGTGCAG CAGCACGGTG ACGTCCCGTT GCCGCTGGAG CCTGCCGATG
 24060

ACGGGGATCA TCGATTCCGC GATCTCCTCG CGGGTCTTCC AGTTGGTGAA CGAGTCCTCG
 24120

TTGACAGTCA CAGATCCATC TTTCGAGCTA GCGGGCGCTC ATATGTAAAC CCGATGGGTG
 24180

TGATCATCTG CCGGGCGCGG GTGTGATGTA CGACCAGCTC GGGGCGGGTG CGGGACGTCG
 24240

GGGATGGGGG CGCGGAGCAC CGGGTGCGGG GTGTGGGTGG CCGGTGGGGG GACGTGCGGC
 24300

GGGGTGCAG TACGTGCGGT GCGGGTGCGG AACGTGCGGC GAGGTGCGGA ACGTGCGGTG
 24360

CGGGTGCGGA ACGTGCGGC
 24379

图 1

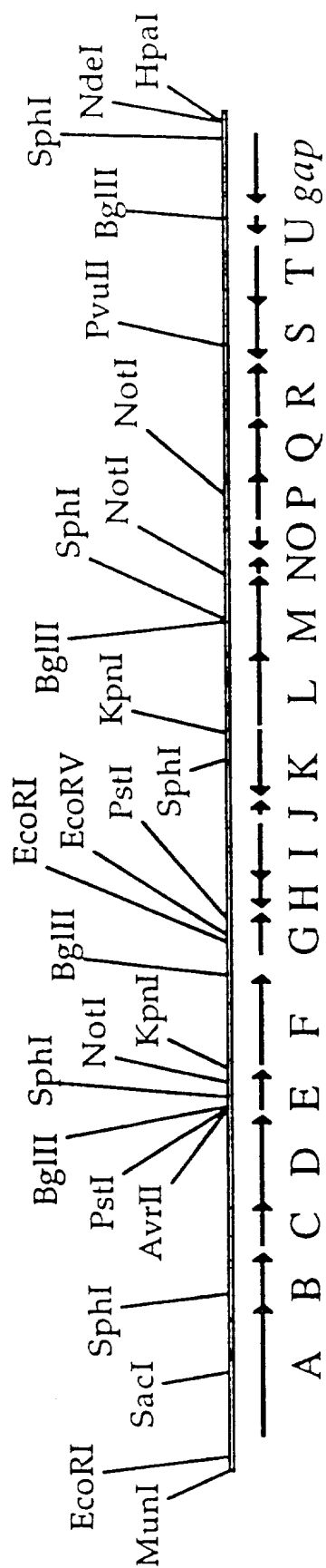


图 2

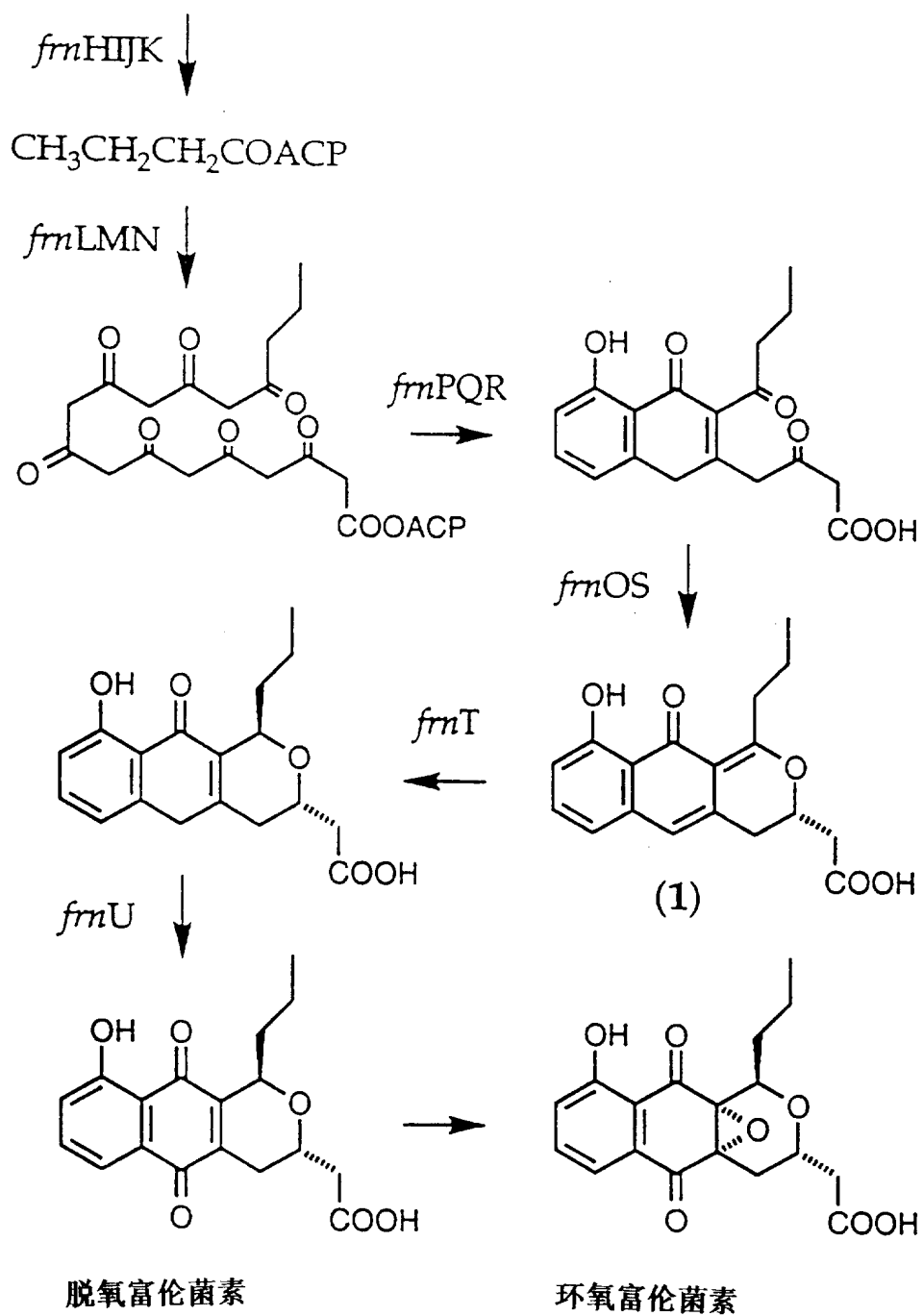


图 3

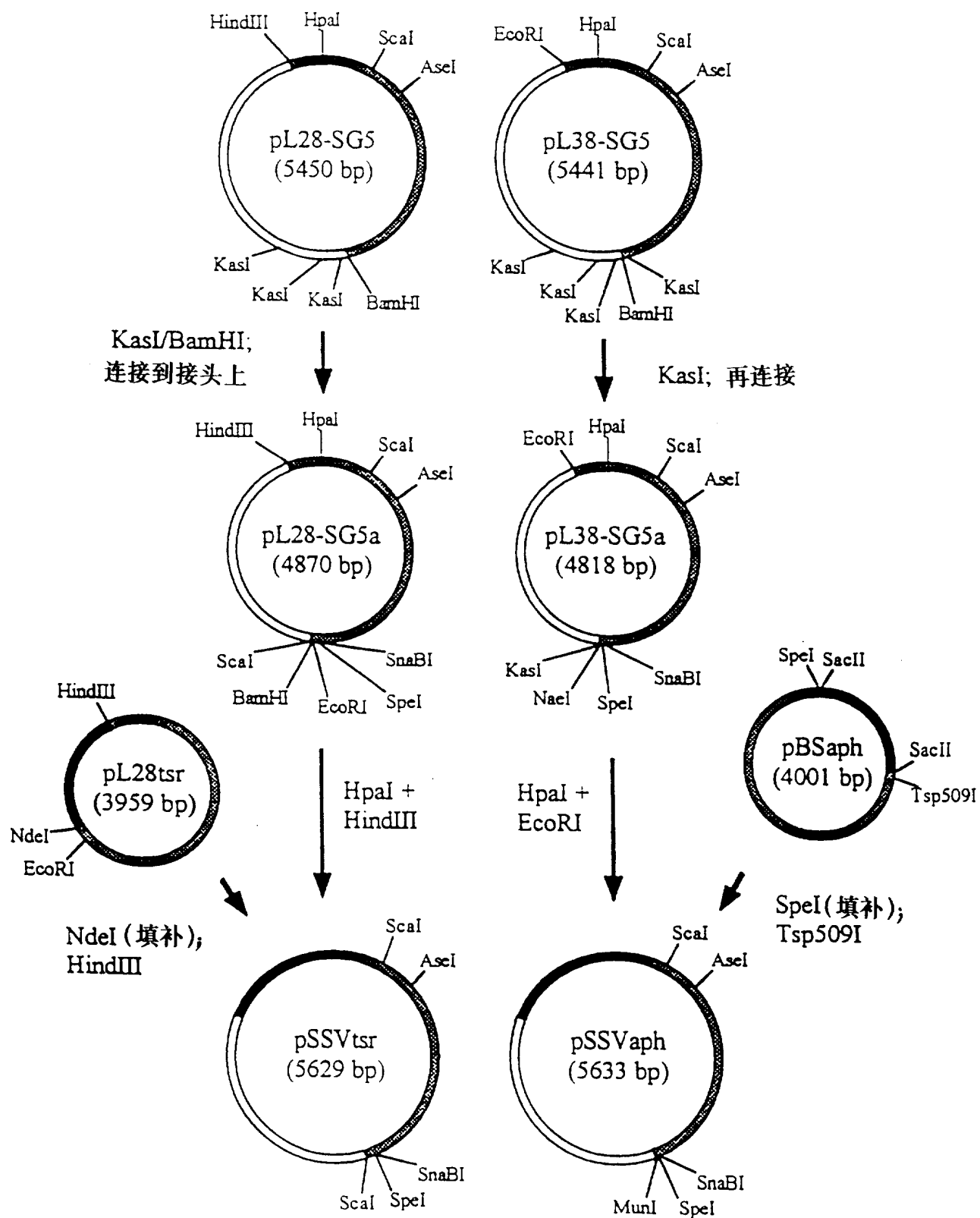


图 4

合成 ermE* 启动子:

BamHI KpnI BglII

GGATCCGGTACCAGATCTGCTTGTGGGGGTTGATGGGGCATGCCAGGGTCCACGGGGTC

GATCTTTCACGGGTGCGGACAGGTGGGGGAGCATCTGACCGACGGGGTCCACAGGTGGC

ACCGGCATGCTGTGTGTCGGCCACAATCGTCCCGGTTGGTAGCATCCagcgata

BamHI

合成 tipA 启动子:

BamHI NdeI KpnI BglII SmaI

GGATCCCATATGCGGTACCAGATCTTCCCGGGGCTGAGGGAGCCCGACCGCCACCGGGGGGCTCA

CGGGGTGGCAACGGGAAGGTCCGGGGCTTGCACCTCACGTACAGTCAGGAGCCAGCGTCCA

CGGGGTGACAGCAAGGGAGCGGCATATG

NdeI

图 5

