

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.



[12] 发明专利申请公开说明书

[21] 申请号 200480025053.3

C12N 15/09 (2006.01)

C12N 1/19 (2006.01)

C12N 1/21 (2006.01)

C12N 5/10 (2006.01)

C12P 21/02 (2006.01)

C12Q 1/02 (2006.01)

[43] 公开日 2006 年 10 月 11 日

[11] 公开号 CN 1845992A

[22] 申请日 2004.8.31

[21] 申请号 200480025053.3

[30] 优先权

[32] 2003.9.1 [33] JP [31] 308773/2003

[86] 国际申请 PCT/JP2004/012523 2004.8.31

[87] 国际公布 WO2005/021747 日 2005.3.10

[85] 进入国家阶段日期 2006.3.1

[71] 申请人 公立大学法人横浜市立大学

地址 日本国神奈川

[72] 发明人 广明秀一 天野刚志 合田名都子

[74] 专利代理机构 中科专利商标代理有限公司

代理人 朱丹

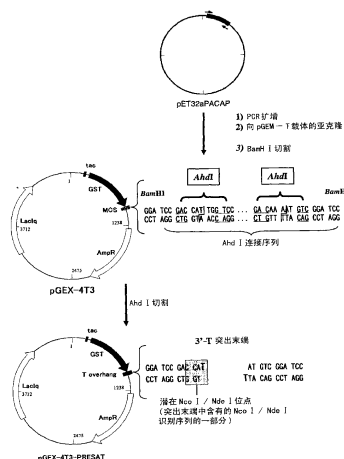
权利要求书 3 页 说明书 64 页 序列表 24 页
附图 7 页

[54] 发明名称

新型载体及其利用

[57] 摘要

本发明提供一种载体，其特征在于，具有被第一限制性酶识别的碱基序列不同的两个切割识别序列，其中一个切割识别序列含有由与第一限制性酶不同的第二限制性酶识别的切割识别序列的一部分，另一个切割识别序列不含由上述第二限制性酶识别的切割识别序列的一部分。本发明还提供利用上述载体来制造蛋白质和/或蛋白质结构域的方法。由此，本发明提供一种用于减少在宿主中重组表达基因得到蛋白质的作业的劳力和时间的工具。



1. 一种载体, 其中,
 - 5 具有被第一限制性酶识别的碱基序列不同的两个切割识别序列, 其中一个切割识别序列含有由不同于第一限制性酶的第二限制性酶识别的切割识别序列的一部分, 而另一个切割识别序列不含由所述第二限制性酶识别的切割识别序列的一部分。
 2. 根据权利要求1所述的载体, 其中,
 - 10 第一限制性酶是在已切割的DNA上产生3'突出末端的限制性酶。
 3. 根据权利要求2所述的载体, 其中,
3'突出末端的碱基为胸腺嘧啶。
 4. 根据权利要求1~3中任意一项所述的载体, 其中,
第一限制性酶是从Ahd I或其同裂酶、Xcm I或其同裂酶、以及其组
 - 15 合中选择的。
 5. 根据权利要求4所述的载体, 其中,
第一限制性酶是从Ahd I或其同裂酶、Xcm I或其同裂酶、以及其组合中选择的, 第二限制性酶是从Nco I、Dsa I、EcoT14 I、Nde I、Sfe I、Ava I、Sml I、Xho I、Pst I、EcoT22 I、PshBI、Acc I、ApaLI、Psi I、Sca I、
 - 20 Spl I、Bsp1107 I、Bsp1286 I、BspT107 I、Bsp1407 I、Tat I、HgiCI、Kpn I、Spe I、Bln I、Nhe I、Xba I、Ssp I、EcoRV以及它们的同裂酶中选择的。
 6. 根据权利要求4所述的载体, 其中,
第一限制性酶是Xcm I或其同裂酶, 第二限制性酶是从AflII I、BspLU11 I、Nsp I、Sph I、BspHI、BsaAI、PmaCI、SnaBI、NspBII、PvuII、BanII、HgiAI、Sac I以及它们的同裂酶中选择的。
 - 25 7. 根据权利要求1~6中任意一项所述的载体, 其中,

将由第一限制性酶识别的碱基序列不同的两个切割识别序列插入到多克隆位点中。

8. 根据权利要求 1~7 中任意一项所述的载体, 其中,

载体含有某种蛋白质的编码区域, 结合对蛋白质和/或蛋白质结构域进行编码的外来基因时, 能够以与所述区域编码的蛋白质结合成融合蛋白质的形式使所述外来基因编码的蛋白质和/或蛋白质结构域表达。

9. 根据权利要求 8 所述的载体, 其中,

在某种蛋白质的编码区域的羧基末端侧, 具有由第一限制性酶识别的碱基序列不同的两个切割识别序列, 其中靠近所述蛋白质的编码区域的一侧的切割识别序列, 含有由与第一限制性酶不同的第二限制性酶所识别的切割识别序列的一部分, 远离所述蛋白质的编码区域的一侧的切割识别序列, 不含由所述第二限制性酶识别的切割识别序列的一部分。

10. 根据权利要求 8 或 9 所述的载体, 其中,

含有谷胱甘肽 S 转移酶的编码区域, 结合对蛋白质和/或蛋白质结构域进行编码的外来基因时, 能够以与谷胱甘肽 S 转移酶结合成融合蛋白质的形式使所述外来基因编码的蛋白质和/或蛋白质结构域表达。

11. 根据权利要求 10 所述的载体, 其中,

谷胱甘肽 S 转移酶来源于哺乳类、日本双盘吸虫或大肠杆菌。

12. 一种开环状载体, 其中,

由用第一限制性酶处理权利要求 1~11 中任意一项所述的载体得到的双链 DNA 构成。

13. 一种重组载体, 其中,

是将含有下述序列的外来基因结合到权利要求 1 所述的载体中而成, 所述的序列被设计成当顺向插入到用第一限制性酶处理过的权利要求 1 所述的载体时不出现第二限制性酶的切割识别序列, 当逆向插入时出现第二限制性酶的切割识别序列。

14. 根据权利要求 13 所述的重组载体, 其中,

外来基因被顺向插入到用第一限制性酶处理过的权利要求 1 所述的载体。

15. 根据权利要求 13 或 14 所述的重组载体, 其中,

外来基因编码蛋白质和/或蛋白质结构域。

16. 一种转化体或其后代, 其中,
含有权利要求 13 所述的重组载体。

17. 一种权利要求 1 所述的载体的制造方法, 其中,
5 包括: 将含有下述两个切割识别序列的双链 DNA 导入到质粒、噬菌体、粘粒、P1 载体、细菌人工染色体载体或酵母人工染色体载体的多克隆位点中, 所述的两个切割识别序列是被第一限制性酶识别的碱基序列不同的两个切割识别序列, 其中一个切割识别序列含有由不同于第一限制性酶的第二限制性酶所识别的切割识别序列的一部分, 另一个切割识别序列
10 不含由上述第二限制性酶识别的切割识别序列的一部分。

18. 一种双链 DNA, 其中,
含有下述两个切割识别序列, 所述的两个切割识别序列是被第一限制性酶识别的碱基序列不同的两个切割识别序列, 其中一个切割识别序列含有由不同于第一限制性酶的第二限制性酶所识别的切割识别序列的一部分,
15 另一个切割识别序列不含由上述第二限制性酶识别的切割识别序列的一部分。

19. 一种方法使用权利要求 1 所述的载体, 制造由外来基因编码的蛋白质和/或蛋白质结构域的方法。

20. 一种使用权利要求 1 所述的载体, 以使外来基因被逆向插入到载体而成的重组载体开环、将外来基因被顺向插入到载体而成的重组载体闭环的状态进行保留的方法。

21. 一种使用权利要求 1 所述的载体, 对由外来基因编码的蛋白质和/或蛋白质结构域的表达量和/或溶解度进行试验的方法。

新型载体及其利用

5

技术领域

本发明涉及一种新型载体及其利用。

背景技术

10 在所谓通常被称作后基因组 (post genome) 研究的结构蛋白质组学 (proteomics)、功能蛋白质组学的学科领域中, 用大肠杆菌同时并行重组不同基因产物并进行表达来获得蛋白质样品的技术是非常重要的。

以往, 为了使用目的基因获得蛋白质样品, 用 PCR 法扩增目的基因, 将 PCR 产物结合到亚克隆用载体中, 由其使宿主转化, 使已转化的宿主
15 形成单菌落, 从各菌落来源的宿主纯化质粒, 通过确定质粒的碱基序列, 来选择已插入目的基因的克隆。然后, 从已选择的克隆, 用限制性酶切下含有目的基因的 DNA 片断, 纯化已切下的 DNA 片断, 再结合到融合蛋白质表达载体中, 由其转化宿主, 使已转化的宿主形成单菌落, 从各菌落来源的宿主纯化质粒, 用已纯化的质粒来转化表达用宿主, 使编码目的基
20 因的蛋白质表达。上述操作最快也需要 9 天时间, 而且需要对用限制性酶从亚克隆用载体切下含目的基因的 DNA 片断的工序以及纯化已切下的 DNA 片断的工序操作熟练。

发明内容

25 本发明的目的在于, 提供一种用于减少在宿主中重组并表达基因以得到蛋白质的作业的劳力和时间的工具。

另外, 本发明的目的还在于, 提供一种利用上述工具、制造由基因编码的蛋白质的方法。

在用于 PCR 法的酶中, 对于最具有代表性的 Taq DNA 聚合酶
30 (polymerase) /Tth DNA 聚合酶等, 已知具有在被复制的 DNA 产物的 3'

末端不依赖于模板而以一个残基附加 dA 的副反应活性。所以，具有在使用 PCR 法扩增的双链 DNA 片断的大部分以一个残基附加 dA 的 3'端突出结构。

在这里，准备质粒载体 (plasmid vector)，所述的质粒载体具有使与 dA 形成互补的碱基对的 dT 在 3'端突出的结构，当与上述 PCR 扩增产物混合时，立即生成利用一个碱基的突出末端的碱基对，这样当混合 DNA 连接酶 (ligase) 时，载体与 PCR 扩增产物发生连接 (ligation) 反应。另一方面，由于 DNA 连接酶不会发生具有错配的 DNA 末端之间的连接反应，所以已知存在极难生成不含有 PCR 扩增产物而只用质粒载体来连接、自身环化的产物这样的现象。

利用该性质，预先作成具有 3'-dT 突出结构的开环载体 (T-载体 (vector))，将不进行利用限制性酶的处理或末端的平滑化处理而在其中混合 PCR 产物并进行连接的方法称为 TA-克隆法。

在 T-载体作成法中，原理上已知有三种。

第一方法为使用 TdT 的方法。TdT (末端脱氧核苷酸转移酶 (terminal deoxynucleotidyltransferase)) 是不依赖模板而向 DNA 末端附加任意脱氧核苷酸 (deoxynucleotide) 的酶。因而，首先用提供平滑末端的限制性酶切割质粒，接着只要在 dTTP 存在下使 TdT 发挥作用，就可以作成在 3'上附加一个或多个 dT 的载体。但是，尚无利用该方法作成 T-载体的报道。

第二方法是使用 Taq DNA 聚合酶的 TdT 活性的方法。如上所述，在 PCR 反应液中，Taq DNA 聚合酶只向 DNA 的平滑末端附加一分子 dA。不过，在 Mg^{2+} 过剩的条件下，在使 dNTP 中只存在 dTTP 的条件下，当使 Taq 作用于 DNA 的平滑末端时，由于向 DNA 附加一个残基的 dT，所以利用该性质制作 T-载体是可能的。因而，只要用提供平滑末端的限制性酶切割质粒，然后在 dTTP、 Mg^{2+} 存在下使 Taq 作用于此处，就能够作成向 3'附加一个 dT 的载体。在该方法中，通过在生成平滑末端时使用的限制性酶的识别序列，显著改变了 T 突出末端的生成效率。在已报道的例子中，几乎都是使用 Eco RV 的例子，还已知使用 Sma I 时效率更差。避免这些情况的方法已在 WWW 上公开。

第三方法是设计具有两个直接生成 T-突出末端的限制性酶位点的“盒 (cassette)” DNA 序列，将其用限制性酶消化的方法。作为将能够生成 3'-T-残基突出末端那样的限制性酶用于实际的 T-载体的例子，分别在论文中报道有 Xcm I、Eam11051、Ahd I 的例子。在该方法中，设计含有两处上述限制性酶位点的“盒” DNA 序列，向目的质粒中插入上述盒之后，用该限制性酶切割上述质粒，此时可以制作两侧成为 dT 突出末端的开环状质粒。（此次本发明人等使用 Ahd I 制作 T-载体。如果该方法，可以设计 (design) 成为“非对称的” dT 突出末端的开环状质粒。另外，在该方法中得到的开环状质粒几乎 100% 会成为 dT 突出末端，质量非常高。）

但是，在以往的 TA-克隆 (cloning) 法中存在缺陷。就是不能控制被插入到载体中的目的基因的方向性。结果，例如到目前为止，可以作为 T-载体制作的质粒几乎都主要只用于 (1) 未知基因克隆 (或者亚克隆 (subcloning)) 用载体、(2) 具有新型启动子活性的基因的克隆用载体这两种，用于将作为蛋白质表达的区域 (ORF) 直接 TA-克隆到表达载体的 T-载体的例子很少。在上述的应用方法中，ORF 的方向性不存在问题。只有在进行蛋白质表达实验的情况下，ORF 的方向性成为问题。（此次，本发明人等为了简单地选择只在顺向插入外来基因 (PCR 产物) 的质粒，设计的序列是只在逆向插入外来基因时才出现限制性酶 Nco I 或 Nde I 的切割序列那样的序列。这样可以选择性地只将表达与谷胱甘肽 S 转移酶 [GST] 等顺向融合的融合蛋白质的质粒转换到大肠杆菌。）

这样，就用于使外来基因在大肠杆菌中表达为 GST 融合蛋白质的载体而言，作为市售品，Amersham Pharmacia Biotech 公司出售有在 *Sjistosoma Japonicum* 来源 GST 的编码区域之后具备蛋白酶识别切割序列、多克隆位点 (multicloning site) 的 pGEX 系列。就 GST 融合蛋白质而言，由于其溶解度较高，在大肠杆菌内的蛋白质表达量良好；只要具有利用谷胱甘肽固定化担体 (珠子 (beads)) 的亲合层析，大致用一步就可以纯化蛋白质；有在谷胱甘肽珠子上固定化目的融合蛋白质的状态下，更加容易地观察与其靶分子 (蛋白质、核酸、低分子等) 之间的相互作用的方法 (pull-down 检测 (assay))；与不带标签 (tag) 的蛋白质相比，当

在利用表面胞质基因（plasmon）共振的分子间相互作用实验的检测芯片上固定化时更难以损坏功能等，从上述几点出发，其是在生化学、分子生物学等中使用非常广泛的表达载体之一。通过将该载体改良到 TA-载体中，对于并行制作表达多个种类或长度的蛋白质或蛋白质片断的表达系统是非常有用的。

在基因组项目（genome project）结束后，为了鉴定功能性蛋白质和蛋白质的功能域并将其用于工业中，其立体结构的确定和分子功能的确定这两方面的高速化开始变得非常重要。为此，需要能够使对可用于一系列研究的活性进行保持的蛋白质样品成为可溶性蛋白质。通常在大肠杆菌中表达人或哺乳类、植物来源的蛋白质的全长或一部分基因时，蛋白质不溶而给结构、功能分析的样品调制带来困难，这样的情况非常多。在这种情况下，已知有如下所述的例子：通过改变作为蛋白质表达的区域、长度（将该方法称为结构域配位（domain matching）或结构域化）、或者一处或多处导入氨基酸的部位特异性突变来使溶解度急剧提高的例子。相同的方法还被用在针对作为可溶性得到的蛋白质而进行的在以其立体结构确定作为目的的 NMR 样品作成或结晶化时的溶解度提高或结晶性改善的过程中。其中，并行取得多种基因片断的过程，通过 PCR 法的进步、和 PCR 引物的化学合成价格的降低，其短时间、低成本成为可能。因而，目前的技术课题是，在制作表达系统时的人、时间成本，和迅速区分目的蛋白质是否为可溶性的方法。将 GST 融合蛋白质用于该目的的优点在于，在通过离心操作将含有 GST 融合蛋白质的大肠杆菌碎裂物分为可溶性、不溶性组分之后，在大肠杆菌蛋白质混合物中，通过将 GST 的酶活性作为指标进行显色反应，在例如 96 孔板上并行且在短时间内比色定量测量 GST 融合蛋白质的量成为可能。

本发明人等在考虑上述情况的基础上，尝试进行了下述试验。

1. 使用利用上述的“Sma I /Taq”的方法尝试制作 GST-T 载体。虽然大致完成了这样的载体，但没有重复性，另外 PCR 产物的亚克隆效率也不好，所以不具有实用性。

2. 由于从 WWW 上的文书获得了用于改善在利用 Sma I /Taq 进行时的效率的改良方法，所以对其进行尝试，但未见改善。

3. 考虑到由 Sma I 切割的平滑末端的碱基序列不是低效率的原因，所以也尝试了在用 Klenow 片断（fragment）对由 BamH I 切割产生的突出末端进行平滑化之后，用 Taq 附加 dT 这样的单独的改良法，但未见改善。

5 4. 战略变更成上述的“使用限制性酶盒（casette）的方法”，作为限制性酶决定使用 Ahd I。

5. 不限于 pGEX-2T/pGEX-4T3（GST 融合载体），当使 Ahd I 作用于具有抗生素耐受性的质粒时，判明其氨苄青霉素（ampicilin）耐受性基因会被切割。所以，制作导入沉默的部位特异性突变而使该 Amp 耐受性基因中含有的 Ahd I 切割部位发生突变的 pGEX-4T3。（将用于导入部位特异性突变的 primer=Anti Ahd I 的序列显示于序列表的序列号 21 中）。

6. 制作 Ahd I 一连接序列（linker）。即，利用序列号 22 和 23 的序列的 PCR 引物，作为模板使用 pET32a（Novagene 公司制）的衍生物 pET32aPACAP，进行 PCR，得到约 500bp 的片断，通过常规方法将其结合到已有的 TA-载体（此次为 Promega 产品 pGEM-T），在确认序列之后，扩增克隆，得到纯化后的质粒，使 BamH I 作用于该质粒，得到 Ahd I 一连接序列（约 500bp）。

7. 制作成为 GST-T 载体（GST-T 载体：GST 融合的 TA 载体的意思）的母体的载体。即，扩增在 5 中得到的 pGEX-4T3 的突变体，用 BamH I 切割，在该 BamH I 位点插入在 6 中得到的 Ahd I 连接序列，通过连接（ligation）使其闭环。分离正确制作的克隆，扩增。其中，原理上在此阶段的 Ahd I 连接序列上不存在方向性（以任何方向进入均可）。

8. 制作 GST-T 载体。即，用 Ahd I 切割 7，用琼脂糖凝胶（agarose gel）分离纯化开环状质粒。该载体由于在 3' 具有 dT 突出末端，所以可以用于 TA-克隆。

9. 在使 Nco I 发挥作用时，逆向插入外来基因的载体（如质粒）被切割，导入了该载体的宿主则不能生育，其结果，制作了可以用于只选择宿主（Nco I ORF 选择法）的 Ahd I 连接序列，其中所述的宿主导入了顺向插入外来基因的载体。制作方法与 6 相同，但在引物中使用了序列号

22 和 24 的序列的组。将已制作的、可以用于 Nco I ORF 选择法的 Ahd I 连接序列的序列显示于序列号 26 和 27 中。具有序列号 26 的序列的 DNA 链与具有序列号 27 的序列的 DNA 链会合形成的双链 DNA 为 Ahd I 连接序列。

- 5 10. 制作可以进行 Nco I ORF 选择法的 dGST-T 载体（方向性（directional）GST-T 载体：可以进行 ORF 的方向选择的 GST-T 载体的意思）。制作方法与上述 7、8 相同。不过，在确认与 7 对应时的序列时，选择在与 GST 的融合部分相当于引物的序列号 22 的序列在 5' 侧（GST 侧）的载体，得到成为可以进行 Nco I ORF 选择法的载体的母体的闭环状载体。另外，该载体只要变更对插入的外来基因进行扩增的 PCR 引物的设计，就会使 Nde I 发挥作用来代替 Nco I，切割以逆向插入外来基因的载体（如质粒），导入该载体的宿主不能生育，其结果，也可以只选择导入了顺向插入外来基因的载体的宿主（Nde I ORF 选择法）。

本发明正是基于以上实验的经过和结果而完成的发明。

- 15 本发明之一提供一种载体，其特征在于，具有被第一限制性酶识别的碱基序列不同的两个切割识别序列，其中一个切割识别序列含有通过与第一限制性酶不同的第二限制性酶进行识别的切割识别序列的一部分，而另一个切割识别序列不含有被上述第二限制性酶识别的切割识别序列的一部分。

- 20 第一限制性酶可以是在已切割的 DNA 上产生 3'突出末端的限制性酶，作为 3'突出末端的碱基，可以例示为胸腺嘧啶。

第一限制性酶可以从 Ahd I 或其同裂酶（isoschizomer）、Xcm I 或其同裂酶（isoschizomer）、以及它们的组合中选择。

- 25 当第一限制性酶为 Ahd I 或其同裂酶时，被 Ahd I 识别的碱基序列不同的两个切割识别序列中的一个，优选下述的序列（I）和（II）表示，另一个优选下述序列（III）和（IV）表示。

5' -GACX^{1a}X^{2a}TX^{3a}X^{4a}GTC-3' (I)

3' -CTGX^{1b}X^{2b}AX^{3b}X^{4b}CAG-5' (II)

5' -GACX^{5a}X^{6a}AX^{7a}X^{8a}GTC-3' (III)

- 30 3' -CTGX^{5b}X^{6b}TX^{7b}X^{8b}CAG-5' (IV)

(在序列(I)~(IV)中, G为鸟嘌呤、A为腺嘌呤、C为胞嘧啶、T为胸腺嘧啶,

$X^{1a} \sim X^{8a}$ 和 $X^{1b} \sim X^{8b}$ 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基, X^{1a} 与 X^{1b} 为彼此可以形成配对的碱基, X^{2a} 与 X^{2b} 为彼此
5 可以形成配对的碱基, X^{3a} 与 X^{3b} 为彼此可以形成配对的碱基, X^{4a} 与 X^{4b} 为彼此可以形成配对的碱基, X^{5a} 与 X^{5b} 为彼此可以形成配对的碱基, X^{6a} 与 X^{6b} 为彼此可以形成配对的碱基, X^{7a} 与 X^{7b} 为彼此可以形成配对的碱基, X^{8a} 与 X^{8b} 为彼此可以形成配对的碱基,

序列 $CX^{1a}X^{2a}T$ 为含有第二限制性酶的切割识别序列的5'侧的部分序
10 列的碱基序列,

序列 $CX^{8b}X^{7b}T$ 为不含在序列 $CX^{1a}X^{2a}T$ 中含有的第二限制性酶的切割识别序列的部分序列的碱基序列,

5'—和3'—表示双链DNA的各自的末端。)

当第一限制性酶为Xcm I或其同裂酶时, 被Xcm I识别的碱基序列
15 不同的两个切割识别序列中的一个, 优选下述的序列(V)和(VI)表示, 另一个优选下述序列(VII)和(VIII)表示。

5' -CCAY^{1a}Y^{2a}Y^{3a}Y^{4a}TY^{5a}Y^{6a}Y^{7a}Y^{8a}TGG-3' (V)

3' -GGTY^{1b}Y^{2b}Y^{3b}Y^{4b}AY^{5b}Y^{6b}Y^{7b}Y^{8b}ACC-5' (VI)

5' -CCAY^{9a}Y^{10a}Y^{11a}Y^{12a}AY^{13a}Y^{14a}Y^{15a}Y^{16a}TGG-3' (VII)

20 3' -GGTY^{9b}Y^{10b}Y^{11b}Y^{12b}TY^{13b}Y^{14b}Y^{15b}Y^{16b}ACC-5' (VIII)

(在序列(V)~(VIII)中, G为鸟嘌呤、A为腺嘌呤、C为胞嘧啶、T为胸腺嘧啶,

$Y^{1a} \sim Y^{16a}$ 和 $Y^{1b} \sim Y^{16b}$ 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基, Y^{1a} 与 Y^{1b} 为彼此可以形成配对的碱基, Y^{2a} 与 Y^{2b} 为彼此
25 可以形成配对的碱基, Y^{3a} 与 Y^{3b} 为彼此可以形成配对的碱基, Y^{4a} 与 Y^{4b} 为彼此可以形成配对的碱基, Y^{5a} 与 Y^{5b} 为彼此可以形成配对的碱基, Y^{6a} 与 Y^{6b} 为彼此可以形成配对的碱基, Y^{7a} 与 Y^{7b} 为彼此可以形成配对的碱基, Y^{8a} 与 Y^{8b} 为彼此可以形成配对的碱基, Y^{9a} 与 Y^{9b} 为彼此可以形成配对的碱基, Y^{10a} 与 Y^{10b} 为彼此可以形成配对的碱基, Y^{11a} 与 Y^{11b} 为彼此可以形

成配对的碱基, Y^{12a} 与 Y^{12b} 为彼此可以形成配对的碱基, Y^{13a} 与 Y^{13b} 为彼此可以形成配对的碱基, Y^{14a} 与 Y^{14b} 为彼此可以形成配对的碱基, Y^{15a} 与 Y^{15b} 为彼此可以形成配对的碱基, Y^{16a} 与 Y^{16b} 为彼此可以形成配对的碱基,

序列 $Y^{1a}Y^{2a}Y^{3a}Y^{4a}T$ 为含有第二限制性酶的切割识别序列的 5' 侧的部分序列的碱基序列,

序列 $Y^{16b}Y^{15b}Y^{14b}Y^{13b}T$ 为不含在序列 $Y^{1a}Y^{2a}Y^{3a}Y^{4a}T$ 中含有的第二限制性酶的切割识别序列的部分序列的碱基序列,

5'—和 3'—表示双链 DNA 的各自的末端。)

当第一限制性酶为 Ahd I 或其同裂酶、与 Xcm I 或其同裂酶的组合时, 被第一限制性酶识别的碱基序列不同的两个切割识别序列中的一个, 是由 Ahd I 或其同裂酶识别的切割识别序列, 而且可以含有由不同于 Ahd I 或其同裂酶的第二限制性酶所识别的切割序列的一部分; 而另一个切割识别序列是由 Xcm I 或其同裂酶识别的切割识别序列, 而且也可以不含由上述第二限制性酶识别的切割识别序列的一部分。是由 Ahd I 或其同裂酶识别的切割识别序列且含有由不同于 Ahd I 或其同裂酶的第二限制性酶识别的切割序列的一部分的切割识别序列, 优选下述的序列 (I) 和 (II) 表示。是由 Xcm I 或其同裂酶识别的切割识别序列且不含由第二限制性酶识别的切割识别序列的一部分的切割识别序列, 优选下述的序列 (VII) 和 (VIII) 表示。

5' -GACX^{1a}X^{2a}TX^{3a}X^{4a}GTC-3' (I)

3' -CTGX^{1b}X^{2b}AX^{3b}X^{4b}CAG-5' (II)

(在序列 (I) ~ (II) 中, G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧啶、T 为胸腺嘧啶,

$X^{1a} \sim X^{4a}$ 和 $X^{1b} \sim X^{4b}$ 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基, X^{1a} 与 X^{1b} 为彼此可以形成配对的碱基, X^{2a} 与 X^{2b} 为彼此可以形成配对的碱基, X^{3a} 与 X^{3b} 为彼此可以形成配对的碱基, X^{4a} 与 X^{4b} 为彼此可以形成配对的碱基,

序列 $CX^{1a}X^{2a}T$ 为含有第二限制性酶的切割识别序列的 5' 侧的部分序列的碱基序列,

5'—和 3'—表示双链 DNA 的各自的末端。)

5' -CCAY^{9a}Y^{10a}Y^{11a}Y^{12a}AY^{13a}Y^{14a}Y^{15a}Y^{16a}TGG-3' (VII)

3' -GGTY^{9b}Y^{10b}Y^{11b}Y^{12b}TY^{13b}Y^{14b}Y^{15b}Y^{16b}ACC-5' (VIII)

(在序列 (VII) ~ (VIII) 中, G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧啶、T 为胸腺嘧啶)

5 Y^{9a}~Y^{16a} 和 Y^{9b}~Y^{16b} 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧的任意碱基, Y^{9a} 与 Y^{9b} 为彼此可以形成配对的碱基, Y^{10a} 与 Y^{10b} 为彼此可以形成配对的碱基, Y^{11a} 与 Y^{11b} 为彼此可以形成配对的碱基, Y^{12a} 与 Y^{12b} 为彼此可以形成配对的碱基, Y^{13a} 与 Y^{13b} 为彼此可以形成配对的碱基, Y^{14a} 与 Y^{14b} 为彼此可以形成配对的碱基, Y^{15a} 与 Y^{15b} 为彼此可以形成配对的碱基, Y^{16a} 与 Y^{16b} 为彼此可以形成配对的碱基,

序列 Y^{16b}Y^{15b}Y^{14b}Y^{13b}T 为不含在序列 CX^{1a}X^{2a}T 中含有的第二限制性酶的切割识别序列的部分序列的碱基序列,

5'—和 3'—表示双链 DNA 的各自的末端。)

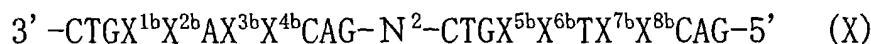
当第一限制性酶是从 Ahd I 或其同裂酶、Xcm I 或其同裂酶、以及其组合中选择时, 第二限制性酶优选从 Nco I、Dsa I、EcoT14 I、Nde I、Sfe I、Ava I、Sml I、Xho I、Pst I、EcoT22 I、PshBI、Acc I、ApaLI、Psi I、Sca I、Spl I、Bsp1107 I、Bsp1286 I、BspT107 I、Bsp1407 I、Tat I、HgiCI、Kpn I、Spe I、Bln I、Nhe I、Xba I、Ssp I、EcoRV 以及它们的同裂酶中选择。

当第二限制性酶为 Xcm I 或其同裂酶时, 第二限制性酶优选从 AflIII、BspLU11 I、Nsp I、Sph I、BspHI、BsaAI、PmaCI、SnaBI、NspBII、PvuII、BanII、HgiAI、Sac I 以及它们的同裂酶中选择。

由第一限制性酶识别的碱基序列不同的两个切割识别序列可以插入到多克隆位点中。

当第一限制性酶为 Ahd I 或其同裂酶时, 优选将用下述序列 (IX) 和 (X) 表示的双链 DNA 插入到多克隆位点中。

30 5' -GACX^{1a}X^{2a}TX^{3a}X^{4a}GTC-N¹-GACX^{5a}X^{6a}AX^{7a}X^{8a}GTC-3' (IX)



(在序列 (IX) 和 (X) 中, G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧啶、T 为胸腺嘧啶,

$\text{X}^{1a} \sim \text{X}^{8a}$ 和 $\text{X}^{1b} \sim \text{X}^{8b}$ 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧
5 的任意碱基, X^{1a} 与 X^{1b} 为彼此可以形成配对的碱基, X^{2a} 与 X^{2b} 为彼此可以形成配对的碱基, X^{3a} 与 X^{3b} 为彼此可以形成配对的碱基, X^{4a} 与 X^{4b} 为彼此可以形成配对的碱基, X^{5a} 与 X^{5b} 为彼此可以形成配对的碱基, X^{6a} 与 X^{6b} 为彼此可以形成配对的碱基, X^{7a} 与 X^{7b} 为彼此可以形成配对的碱基, X^{8a} 与 X^{8b} 为彼此可以形成配对的碱基,

10 序列 $\text{CX}^{1a}\text{X}^{2a}\text{T}$ 为含有第二限制性酶的切割识别序列的 5' 侧的部分序列的碱基序列,

序列 $\text{CX}^{8b}\text{X}^{7b}\text{T}$ 为不含在序列 $\text{CX}^{1a}\text{X}^{2a}\text{T}$ 中含有的第二限制性酶的切割识别序列的部分序列的碱基序列,

5'-和 3'-表示双链 DNA 的各自的末端,

15 N^1 可以不存在, 但当存在时, 为 1~1000 碱基的任意核苷酸序列,

N^2 可以不存在, 但当存在时, 为与 N^1 的核苷酸序列互补的核苷酸序列。)

当第一限制性酶为 Xcm I 或其同裂酶时, 优选将用下述序列 (XI) 和 (XII) 表示的双链 DNA 插入到多克隆位点中。

20 $5' - \text{CCAY}^{1a}\text{Y}^{2a}\text{Y}^{3a}\text{Y}^{4a}\text{TY}^{5a}\text{Y}^{6a}\text{Y}^{7a}\text{Y}^{8a}\text{TGG} - \text{N}^1 - \text{CCAY}^{9a}\text{Y}^{10a}\text{Y}^{11a}\text{Y}^{12a}\text{AY}^{13a}\text{Y}^{14a}\text{Y}^{15a}\text{Y}^{16a}$
 $\text{TGG} - 3' \quad (\text{XI})$

$3' - \text{GGTY}^{1b}\text{Y}^{2b}\text{Y}^{3b}\text{Y}^{4b}\text{AY}^{5b}\text{Y}^{6b}\text{Y}^{7b}\text{Y}^{8b}\text{ACC} - \text{N}^2 - \text{GGTY}^{9b}\text{Y}^{10b}\text{Y}^{11b}\text{Y}^{12b}\text{TY}^{13b}\text{Y}^{14b}\text{Y}^{15b}\text{Y}^{16b}$
 $\text{ACC} - 5' \quad (\text{XII})$

(在序列 (XI) 和 (XII) 中, G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧
25 啶、T 为胸腺嘧啶,

$\text{Y}^{1a} \sim \text{Y}^{16a}$ 和 $\text{Y}^{1b} \sim \text{Y}^{16b}$ 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧的任意碱基, Y^{1a} 与 Y^{1b} 为彼此可以形成配对的碱基, Y^{2a} 与 Y^{2b} 为彼此可以形成配对的碱基, Y^{3a} 与 Y^{3b} 为彼此可以形成配对的碱基, Y^{4a} 与 Y^{4b} 为彼此可以形成配对的碱基, Y^{5a} 与 Y^{5b} 为彼此可以形成配对的碱基, Y^{6a}

与 Y^{6b} 为彼此可以形成配对的碱基, Y^{7a} 与 Y^{7b} 为彼此可以形成配对的碱基, Y^{8a} 与 Y^{8b} 为彼此可以形成配对的碱基, Y^{9a} 与 Y^{9b} 为彼此可以形成配对的碱基, Y^{10a} 与 Y^{10b} 为彼此可以形成配对的碱基, Y^{11a} 与 Y^{11b} 为彼此可以形成配对的碱基, Y^{12a} 与 Y^{12b} 为彼此可以形成配对的碱基, Y^{13a} 与 Y^{13b} 为彼此可以形成配对的碱基, Y^{14a} 与 Y^{14b} 为彼此可以形成配对的碱基, Y^{15a} 与 Y^{15b} 为彼此可以形成配对的碱基, Y^{16a} 与 Y^{16b} 为彼此可以形成配对的碱基,

序列 $Y^{1a}Y^{2a}Y^{3a}Y^{4a}T$ 为含有第二限制性酶的切割识别序列的 5' 侧的部分序列的碱基序列,

序列 $Y^{16b}Y^{15b}Y^{14b}Y^{13b}T$ 为不含在序列 $Y^{1a}Y^{2a}Y^{3a}Y^{4a}T$ 中含有的第二限制性酶的切割识别序列的部分序列的碱基序列,

5'—和 3'—表示双链 DNA 的各自的末端,

N^1 可以不存在, 但当存在时, 为 1~1000 碱基的任意核苷酸序列,

N^2 可以不存在, 但当存在时, 为与 N^1 的核苷酸序列互补的核苷酸序列。)

当第一限制性酶为 Ahd I 或其同裂酶、与 Xcm I 或其同裂酶的组合时, 优选将用下述序列 (XIII) 和 (XIV) 表示的双链 DNA 插入到多克隆位点中。

5' -GACX^{1a}X^{2a}TX^{3a}X^{4a}GTC- N^1 -CCAY^{9a}Y^{10a}Y^{11a}Y^{12a}AY^{13a}Y^{14a}Y^{15a}Y^{16a}TGG-3' (XIII)

3' -CTGX^{1b}X^{2b}AX^{3b}X^{4b}CAG- N^2 -GGTY^{9b}Y^{10b}Y^{11b}Y^{12b}TY^{13b}Y^{14b}Y^{15b}Y^{16b}ACC-5' (XIV)

(在序列 (XIII) 和 (XIV) 中, G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧啶、T 为胸腺嘧啶,

$X^{1a} \sim X^{4a}$ 和 $X^{1b} \sim X^{4b}$ 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基, X^{1a} 与 X^{1b} 为彼此可以形成配对的碱基, X^{2a} 与 X^{2b} 为彼此可以形成配对的碱基, X^{3a} 与 X^{3b} 为彼此可以形成配对的碱基, X^{4a} 与 X^{4b} 为彼此可以形成配对的碱基,

$Y^{9a} \sim Y^{16a}$ 和 $Y^{9b} \sim Y^{16b}$ 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基, Y^{9a} 与 Y^{9b} 为彼此可以形成配对的碱基, Y^{10a} 与 Y^{10b} 为彼此可以形成配对的碱基, Y^{11a} 与 Y^{11b} 为彼此可以形成配对的碱基, Y^{12a} 与 Y^{12b} 为彼此可以形成配对的碱基, Y^{13a} 与 Y^{13b} 为彼此可以形成配对的碱基, Y^{14a} 与 Y^{14b} 为彼此可以形成配对的碱基, Y^{15a} 与 Y^{15b} 为彼此可以形成配对

的碱基, Y^{16a} 与 Y^{16b} 为彼此可以形成配对的碱基,

序列 $CX^{1a}X^{2a}T$ 为含有第二限制性酶的切割识别序列的 5' 侧的部分序列的碱基序列,

5 序列 $Y^{16b}Y^{15b}Y^{14b}Y^{13b}T$ 为不含在序列 $CX^{1a}X^{2a}T$ 中含有的第二限制性酶的切割识别序列的部分序列的碱基序列,

5'—和 3'—表示双链 DNA 的各自的末端。

N^1 可以不存在, 但当存在时, 为 1~1000 碱基的任意核苷酸序列,

N^2 可以不存在, 但当存在时, 为与 N^1 的核苷酸序列互补的核苷酸序列。)

10 在序列 (IX) ~ (XIV) 中, N^1 例如可以是来源于向市售的载体 (如 pET21b、pET32a 等) 的适当的限制性酶位点中插入基因 (例如硫氧还蛋白基因) 和人工基因 (例如 PACAP 等) 而作成的衍生物的序列。优选在 N^1 序列的一部分中含有第二限制性酶位点。在这种情况下, 即使当通过第一限制性酶处理本发明的载体 (下面也称为 “PRESAT—载体”) 所得到的开环状载体 (下面也称为 “开环状 PRESAT—载体”) 的纯化不完全时, 从此处由自身连接产生的 “空的” PRESAT—载体, 由于通过第二限制性酶消化而被切割, 所以在 ORF 选择后不会产生菌落。因而, 其优点在于, 即使开环状 PRESAT—载体的纯化不完全, 本底 (background) 也不会上升。

20 作为 N^1 的一个例子, 可以举出以下的序列。

5' -CACCTGACTGACGACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATCCTCGTC
GATTTCTGGGCAGAGTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCT
GACGAATATCAGGGCAAACCTGACCGTTGCAAACTGAACATCGATCAAAACCCTGGCACTGCG
CCGAAATATGGCATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCA
25 ACCAAAGTGGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGT
TCTGGTTCTGGCCATATGGCTAGCCATCACCACCACCACAGCAGCGGCATTGACGGCCGG
CATAGCGATGGCATCTTTACCGATAGCTATAGCCGCTATCGCAAACAGATGGCGGTGAAAAAG
TATCTGGCGGCGGTGCTGGGCTAATAA-3' (序列号 55)

30 序列号 55 的序列是来源于向 pET21b 的 Nde I 位点插入含有 pET32a 来源的硫氧还蛋白基因的 Nde I 片断、和向该 pET21b 的 Nhe I 位点插入

PACAP 人工基因而作成的 pET32a 的衍生物 pET32PACAP 的序列。

在序列 (IX) ~ (XIV) 中, N^2 是与 N^1 的核苷酸序列互补的核苷酸序列。作为 N^2 的一个例子, 可以举出以下的序列。

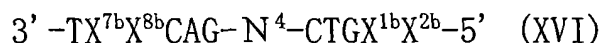
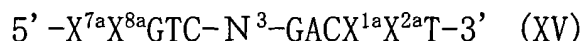
3' -GTGGACTGACTGCTGTCAAACTGTGCCTACATGAGTTTCGCCTGCCCCGCTAGGAGCAG
5 CTAAAGACCCGTCTCACCACGCCAGGCACGTTTTACTAGCGGGGCTAAGACCTACTTTAGCGA
CTGCTTATAGTCCCGTTTTGACTGGCAACGTTTTGACTTGTAGCTAGTTTTGGGACCGTGACGC
GGCTTTATACCGTAGGCACCATAGGGCTGAGACGACGACAAGTTTTTGCCACTTCACCGCCGT
TGGTTTCACCCACGTGACAGATTTCCAGTCAACTTTCTCAAGGAGCTGCGATTGGACCGGCCA
AGACCAAGACCGGTATACCGATCGGTAGTGGTGGTGGTGGTGTCTCGCCGTAAGTCCCGGCC
10 GTATCGCTACCGTAGAAATGGCTATCGATATCGGCGATAGCGTTTGTCTACCGCCACTTTTTTC
ATAGACCGCCGCCACGACCCGATTATT-5' (序列号 56)

序列号 56 的序列是与序列号 55 的序列互补的序列。

本发明之一的载体含有某种蛋白质(如哺乳类、日本双盘吸虫(ジストマジャポニクス)或大肠杆菌来源的谷胱甘肽 S 转移酶(GST)、硫氧
15 还蛋白(TRX)、His-His-His-His-His-His(His6)、镧系元素结合
标签(lanthanide binding tag)(LBT)、Asp-Tyr-Lys-Asp-Asp-Asp-
-Asp-Lys(FLAG(注册商标))、增强绿色荧光蛋白(enhanced green
fluorescence protein)(EGFP)等)的编码区域, 在结合对蛋白质和/或蛋
白质结构域进行编码的外来基因时, 也能够以与上述区域所编码的蛋白质
20 结合而成的融合蛋白质的方式使上述外来基因所编码的蛋白质和/或蛋
白质结构域进行表达。在这种情况下, 在某蛋白质的编码区域的羧基末端侧
具有由第一限制性酶识别的碱基序列不同的两个的切割识别序列, 其中,
靠近上述蛋白质的编码区域的一侧的切割识别序列, 含有由不同于第一限
制性酶的第二限制性酶识别的切割识别序列的一部分, 当远离上述蛋白质的
25 的编码区域的一侧的切割识别序列, 可以不含由上述第二限制性酶识别的
切割识别序列的一部分。

本发明之二提供一种由通过第一限制性酶处理本发明之一的载体而得到的双链 DNA 构成的开环状载体。

作为本发明的开环状载体, 可以例示由用下述序列 (XV) 和 (XVI)
30 表示的双链 DNA 构成的开环状载体。



(在序列 (XV) 和 (XVI) 中, G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧啶、T 为胸腺嘧啶,

5 X^{7a} 、 X^{8a} 、 X^{7b} 和 X^{8b} 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基, X^{7a} 与 X^{7b} 为彼此可以形成配对的碱基, X^{8a} 与 X^{8b} 为彼此可以形成配对的碱基,

X^{1a} 、 X^{2a} 、 X^{1b} 和 X^{2b} 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基, X^{1a} 与 X^{1b} 为彼此可以形成配对的碱基, X^{2a} 与 X^{2b} 为彼此可以形成配对的碱基,

15 序列 $CX^{1a}X^{2a}T$ 是含有 N c o I、D s a I、E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V 以及它们的同裂酶的切割识别序列的 5' 侧的部分序列的碱基序列,

20 序列 $CX^{8b}X^{7b}T$ 为不含在序列 $CX^{1a}X^{2a}T$ 中含有的 N c o I、D s a I、E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V 以及它们的同裂酶的切割识别序列的部分序列的碱基序列,

25 5' 和 3' 表示双链 DNA 的各自的末端,

N^3 为 2000~7000 碱基的任意核苷酸序列, N^4 为与 N^3 的核苷酸序列互补的核苷酸序列。)

在序列 (XV) 和 (XVI) 中, 优选 X^{7a} 为 A、 X^{8a} 为 T、 X^{1a} 为 C、 X^{2a} 为 A、 X^{7b} 为 T、 X^{8b} 为 A、 X^{1b} 为 G、 X^{2b} 为 T。

30 另外, 作为本发明的开环状载体, 可以例示为用下述序列 (XVII) 和

(XVIII) 表示的双链 DNA 构成的开环状载体。

5' -Y^{13a}Y^{14a}Y^{15a}Y^{16a}TGG-N³-CCAY^{1a}Y^{2a}Y^{3a}Y^{4a}T-3' (XVII)

3' -TY^{13b}Y^{14b}Y^{15b}Y^{16b}ACC-N⁴-GGTY^{1b}Y^{2b}Y^{3b}Y^{4b}-5' (XVIII)

(在序列 (XVII) 和 (XVIII) 中, G 为鸟嘌呤、腺嘌呤、胞嘧啶或
5 胸腺嘧,

Y^{1a}、Y^{2a}、Y^{3a}、Y^{4a}、Y^{1b}、Y^{2b}、Y^{3b} 以及 Y^{4b} 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧的任意碱基, Y^{1a} 与 Y^{1b} 为彼此可以形成配对的碱基, Y^{2a} 与 Y^{2b} 为彼此可以形成配对的碱基, Y^{3a} 与 Y^{3b} 为彼此可以形成配对的碱基, Y^{4a} 与 Y^{4b} 为彼此可以形成配对的碱基,

10 序列 Y^{1a}Y^{2a}Y^{3a}Y^{4a}T 为含有 N c o I、D s a I、E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V, A f l I I I、B s p L U 1 1 I、N s p I、S p h I、B s p H I、B s a A I、P m a C I、S n a B I、N s p B I I、P v u I I、B a n I I、H g i A I、S a c I 以及它们的同裂酶的切割识别序列的 5' 侧的部分序列的碱基序列,

20 Y^{13a}、Y^{14a}、Y^{15a}、Y^{16a}、Y^{13b}、Y^{14b}、Y^{15b} 以及 Y^{16b} 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧的任意碱基 Y^{13a} 与 Y^{13b} 为彼此可以形成配对的碱基, Y^{14a} 与 Y^{14b} 为彼此可以形成配对的碱基, Y^{15a} 与 Y^{15b} 为彼此可以形成配对的碱基, Y^{16a} 与 Y^{16b} 为彼此可以形成配对的碱基,

25 序列 Y^{16b}Y^{15b}Y^{14b}Y^{13b}T 为不含在序列 Y^{1a}Y^{2a}Y^{3a}Y^{4a}T 中含有的 N c o I、D s a I、E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V, A f l I I I、B s p L U 1 1 I、N s p I、S p h
30

I、BspHI、BsaAI、PmaCI、SnaBI、NspBII、PvuII、BamII、HgiAI、SacI以及它们的同裂酶的切割识别序列的部分序列的碱基序列，

5'—和3'—表示双链DNA的各自的末端，

5 N³为2000~7000碱基的任意核苷酸序列，N⁴为与N³的核苷酸序列互补的核苷酸序列。)

在序列(XVII)和(XVIII)中，优选Y^{2a}为C、Y^{3a}为C、Y^{4a}为A、Y^{2b}为G、Y^{3b}为G、Y^{4b}为T、Y^{13a}为T、Y^{14b}为G、Y^{15a}为T、Y^{13b}为A、Y^{14b}为C、Y^{15b}为A。

10 进而，作为本发明的开环状载体，可以例示为由下述序列(XIX)和(XX)表示的双链DNA构成的开环状载体。

5' -Y^{13a}Y^{14a}Y^{15a}Y^{16a}TGG-N³-GACX^{1a}X^{2a}T-3' (XIX)

3' -TY^{13b}Y^{14b}Y^{15b}Y^{16b}ACC-5' N⁴-CTGX^{1b}X^{2b}-5' (XX)

15 (在序列(XIX)和(XX)中，G为鸟嘌呤、A为腺嘌呤、C为胞嘧啶、T为胸腺嘧啶，

Y^{13a}、Y^{14a}、Y^{15a}、Y^{16a}、Y^{13b}、Y^{14b}、Y^{15b}以及Y^{16b}分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基，Y^{13a}与Y^{13b}为彼此可以形成配对的碱基，Y^{14a}与Y^{14b}为彼此可以形成配对的碱基，Y^{15a}与Y^{15b}为彼此可以形成配对的碱基，Y^{16a}与Y^{16b}为彼此可以形成配对的碱基，

20 X^{1a}、X^{2a}、X^{1b}和X^{2b}分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基，X^{1a}与X^{1b}为彼此可以形成配对的碱基，X^{2a}与X^{2b}为彼此可以形成配对的碱基，

25 序列CX^{1a}X^{2a}T为含有NcoI、DsaI、EcoT14I、NdeI、SfeI、AvaI、SmlI、XhoI、PstI、EcoT22I、PshBI、AccI、ApaLI、PsiI、ScaI、SplI、Bsp1107I、Bsp1286I、BspT107I、Bsp1407I、TatI、HgiCI、KpnI、SpeI、BlnI、NheI、XbaI、SspI、EcoRV以及它们的同裂酶的切割识别序列的5'侧的部分序列的碱基序列，

30 序列Y^{14b}Y^{13b}T为不含在序列CX^{1a}X^{2a}T中含有的NcoI、DsaI、

E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、
P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s
i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B
s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、
5 S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V以及
它们的同裂酶的切割识别序列的部分序列的碱基序列，

5'—和 3'—表示双链 DNA 的各自的末端，

N³ 为 2000~7000 碱基的任意核苷酸序列，N⁴ 为与 N³ 的核苷酸序列
互补的核苷酸序列。)

10 在序列 (XIX) 和 (XX) 中，优选 X^{1a} 为 C、X^{2a} 为 A、X^{1b} 为 G、X^{2b}
为 T、Y^{13a} 为 T、Y^{14a} 为 G、Y^{15a} 为 T、Y^{13b} 为 A、Y^{14b} 为 C、Y^{15b} 为 A。

在序列 (XV) 和 (XX) 中，N³ 例如可以是来源于市售的载体（例如，
pGEX-4T3、pGEX-2T、pGEX-3X、pET21b、pET32a 等）的序列。
作为 N³ 的一个例子，可以举出以下序列。

15 5' -GGATCCCCGAATTCCTCGGGTCTGACTCGAGCGGCCGCATCGTGACTGACTGACGATCTGCC
TCGCGCGTTCGGTGATGACGGTGAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAG
CTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCG
GGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGATAATTCTTGAAGAC
GAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGA
20 CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATAC
ATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAA
GGAAGAGTATGAGTATTCAACATTTCCGTGTGCGCCTTATCCCTTTTTTGCGGCATTTTGCC
TTCCTGTTTTTGCTCACCCAGAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTG
CACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCG
25 AAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTG
TTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGT
ACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTG
CCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGG
AGCTAACCGCTTTTTTGACAACATGGGGGATCATGTAACCTGCCTTGATCGTTGGGAACCGG
30 AGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAGATGCCTGCAGCAATGGCAACAA

CGTTGCGCAAACCTATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACT
GGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTA
TTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAG
ATGGTAAGCCCTCCCGTATCGTAGTTATCTACACCACGGGGAGCCAGGCAACTATGGATGAAC
5 GAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACGTGCAGACCAAG
TTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTTAATTTAAAGGATCTAGGTGA
AGATCCTTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT
CAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCT
GCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGTGCCGGATCAAGAGCTACCAA
10 CTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGT
AGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAA
TCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGAC
GATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCT
TGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
15 TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCA
CGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCT
GACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCA
ACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGT
TATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCA
20 GCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATT
TTCTCCTTACGCATCTGTGCGGTATTTACACCGCATAAATTCGACACCATCGAATGGTGCA
AAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGA
AACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCG
TGGTGAACCAGGCCAGCCACGTTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGG
25 AGCTGAATTACATTCCCAACCGCGTGGCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTG
GCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTC
GCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTGATGGTAGAACGAAGCGGCGTCGAAGCCT
GTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACATATCCGC
TGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTG
30 ATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGG

GCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTT
CTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGC
CGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGC
TGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAA
5 TGGCGGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACG
ATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGC
TGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATC
AGCTGTTGCCCCTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCT
CTCCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCG
10 GGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACAC
TTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAAC
AGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCC
TGGCGTTACCCAACTTAATCGCCTTGACGACATCCCCCTTTCGCCAGCTGGCGTAATAGCGA
AGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGC
15 CTGGTTTCCGGCACCAGAAGCGGTGCCGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGA
TACTGTCGTCGTCCCTCAAACCTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGT
AACCTATCCCATTACGGTCAATCCGCCGTTTGTTCACGGAGAATCCGACGGGTGTACTC
GCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGG
CGTTGGAATTACGTTATCGACTGCACGGTGACCAATGCTTCTGGCGTCAGGCAGCCATCGGA
20 AGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGGCGCACTC
CCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAG
CTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACA
CAGGAAACAGTATTCATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTCGAACCC
ACTCGACTTCTTTTGAATATCTTGAAGAAAAATATGAAGAGCATTGTATGAGCGCGATGAA
25 GGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTAT
ATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACTGTTATATAGCTGACAAGCAC
AACATGTTGGGTGGTTGTCCAAAAGAGCGTGACAGAGATTTCAATGCTTGAAGGAGCGGTTTTG
GATATTAGATACGGTGTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGAT
TTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCTGAAGATCGTTTATGTCATAAAACATAT
30 TTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTA

TACATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAACGTATTGAA
GCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGC
TGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGTTCCGCGTGGATCC
-3' (序列号 57)

5 序列号 57 的序列是来源于 pGEX-4T3 的序列。

在序列 (XV) 和 (XX) 中, N^4 是与 N^3 的核苷酸序列互补的核苷酸序列。作为 N^4 的一个例子, 可以举出以下序列。

3'-CCTAGGGGCTTAAGGGCCAGCTGAGCTCGCCGGCGTAGCACTGACTGACTGCTAGACGG
AGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTC
10 GAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGC
CCACAGCCCCGCGTCGGTACTGGGTCAGTGCATCGCTATCGCCTCACATATTAAGAACTTCTG
CTTTCCCGGAGCACTATGCGGATAAAAATATCCAATTACAGTACTATTATTACCAAAGAATCT
GCAGTCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATG
TAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTT
15 CCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGAATAAGGGAAAAACGCCGTAAACGG
AAGGACAAAAACGAGTGGGTCTTTGCGACCACTTTCATTTTCTACGACTTCTAGTCAACCCAC
GTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAAAGCGGGGC
TTCTTGCAAAAGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCAC
AACTGCGGCCCCGTTCTCGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCA
20 TGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGAC
GGTATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCC
TCGATTGGCGAAAAACGTGTTGTACCCCTAGTACATTGAGCGGAACTAGCAACCCTTGGCC
TCGACTTACTTCGGTATGGTTTGCTGCTCGCACTGTGGTGCTACGGACGTCGTTACCGTTGTT
GCAACGCGTTTGATAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGA
25 CCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCCGGAAGGCCGACCGACCAAAT
AACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTC
TACCATTGCGGAGGGCATAGCATCAATAGATGTGGTGCCCCCTCGGTCCGTTGATACCTACTTG
CTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGTTC
AAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCTAGATCCACT
30 TCTAGGAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAGCAAGGTGACTCGCA

GTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAGACGCGCATTAGACGA
CGAACGTTTGT TTTT TGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTT
GAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACA
TCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATT
5 AGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTG
CTATCAATGGCCTATTCCGCGTCGCCAGCCGACTTGCCCCCAAGCACGTGTGTGGGTCTGA
ACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCG
AAGGGCTTCCCTCTTCCGCTGTCCATAGGCCATTGCGCGTCCCAGCCTTGTCTCTCGCGT
GCTCCCTCGAAGGTCCCCCTTTCGCGACCATAGAAATATCAGGACAGCCAAAGCGGTGGAGA
10 CTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGCGTCTG
TGCGCCGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCA
ATAGGGGACTAAGACACCTATTGGCATAATGGCGGAACTCACTCGACTATGGCGAGCGGCGT
CGGCTTGCTGGCTCGCGTCGCTCAGTCACTCGCTCCTTCGCCTTCTCGCGGACTACGCCATAA
AAGAGGAATGCGTAGACACGCCATAAAGTGTGGCGTATTTAAGGCTGTGGTAGCTTACCACGT
15 TTTGGAAAGCGCCATACCGTACTATCGCGGGCCTTCTCTCAGTTAAGTCCCACCACTTACACT
TTGGTCATTGCAATATGCTACAGCGTCTCATACGGCCACAGAGAATAGTCTGGCAAAGGGCGC
ACCACTTGGTCCGGTCCGTGCAAAGACGCTTTTTCGCGCCTTTTTCACCTTCGCCGCTACCGCC
TCGACTTAATGTAAGGGTTGGCGCACCGTGTTGTTGACCGCCCGTTTGTCTAGCAACGACTAAC
CGCAACGGTGGAGGTCAGACCGGGACGTGCGCGGCAGCGTTTAAAGCGCCGCTAATTTAGAG
20 CGCGGCTAGTTGACCCACGGTCGCACCACCACAGCTACCATCTTGCTTCGCCGCAGCTTCGGA
CATTTCCGCCGCCACGTGTTAGAAGAGCGCGTTGCGCAGTCACCCGACTAGTAATTGATAGGCG
ACCTACTGGTCCTACGGTAACGACACCTTCGACGGACGTGATTACAAGGCCGCAATAAAGAAC
TACAGAGACTGGTCTGTGGGTAGTTGTCATAATAAAGAGGGTACTTCTGCCATGCGCTGACC
CGCACCTCGTAGACCAGCGTAACCCAGTGGTCGTTTAGCGCGACAATCGCCCGGGTAATTCAA
25 GACAGAGCCGCGCAGACGCAGACCGACCGACCGTATTTATAGAGTGAGCGTTAGTTTAAGTCG
GCTATCGCCTTGCCCTTCCGCTGACCTCACGGTACAGGCCAAAAGTTGTTTGGTACGTTTACG
ACTTACTCCCGTAGCAAGGGTGACGCTACGACCAACGGTTGCTAGTCTACCGCGACCCGCGTT
ACGCGCGGTAATGGCTCAGGCCCGACGCGCAACCAACGCCTATAGAGCCATCACCTATGCTGC
TATGGCTTCTGTCTGAGTACAATATAGGGCGGCAATTGGTGGTAGTTTGTCTAAAAGCGGACG
30 ACCCGTTTGGTCGCACCTGGCGAACGACGTTGAGAGAGTCCCGGTCCGCCACTTCCCGTTAG

TCGACAACGGGCAGAGTGACCACTTTTCTTTTGGTGGGACCGCGGGTTATGCGTTTGGCGGA
GAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAAAGGGCTGACCTTTCGC
CCGTCACCTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTG
AAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTTG
5 TCGATACTGGTACTAATGCCTAAGTGACCGGCAGCAAAATGTTGCAGCACTGACCCTTTTGGG
ACCGCAATGGGTGAATTAGCGGAACGTCGTGTAGGGGGAAAGCGGTCGACCGCATTATCGCT
TCTCCGGGCGTGGCTAGCGGGAAGGGTTGTCAACGCGTCGGACTTACCGCTTACCGCGAAACG
GACCAAAGGCCGTGGTCTTCGCCACGGCCTTTCGACCGACCTCACGCTAGAAGGACTCCGGCT
ATGACAGCAGCAGGGGAGTTTGACCGTCTACGTGCCAATGCTACGCGGGTAGATGTGGTTGCA
10 TTGGATAGGGTAATGCCAGTTAGGCGGCAACAAGGGTGCCTCTTAGGCTGCCCAACAATGAG
CGAGTGTAATTACAATACTTTTCGACCGATGTCCTTCCGGTCTGCGCTTAATAAAAACTACC
GCAACCTTAATGCAATAGCTGACGTGCCACGTGGTTACGAAGACCGCAGTCCGTCCGTAGCCT
TCGACACCATACCGACACGTCCAGCATTTAGTGACGTATTAAGCACAGCGAGTTCGCGTGAG
GGCAAGACCTATTACAAAAACGCGGCTGTAGTATTGCCAAGACCGTTTATAAGACTTTACTC
15 GACAACCTGTTAATTAGTAGCCGAGCATATTACACACCTTAACACTCGCCTATTGTTAAAGTGT
GTCCTTTGTCATAAGTACAGGGGATATGATCCAATAACCTTTTAATCCCGGAACACGTTGGG
TGAGCTGAAGAAAACCTTATAGAACTTCTTTTTATACTTCTCGTAAACATACTCGCGCTACTT
CCACTATTTACCGCTTTGTTTTTCAAACCTTAACCCAAACCTCAAAGGGTTAGAAGGAATAATA
TAACTACCACTACAATTTAATTGTGTCAGATACCGGTAGTATGCAATATATCGACTGTTTCGTG
20 TTGTACAACCCACCAACAGGTTTTCTCGCACGTCTCTAAAGTTACGAACCTCCTCGCCAAAAC
CTATAATCTATGCCACAAAGCTCTTAACGTATATCATTTCTGAACTTTGAGAGTTTCAACTA
AAAGAATCGTTCGATGGACTTTACGACTTTTACAAGCTTCTAGCAAATACAGTATTTTGTATA
AATTTACCACTAGTACATTGGGTAGGACTGAAGTACAACATACTGCGAGAACTACAACAAAAT
ATGTACCTGGGTACACGGACCTACGCAAGGGTTTTAATCAAACAAAATTTTTTGCATAACTT
25 CGATAGGGTGTTTAACTATTCATGAACTTTAGGTCGTTTCATATATCGTACCGGAAACGTCCCG
ACCGTTCGGTGCAAACCACCGCTGGTAGGAGGTTTTAGCCTAGACCAAGGCGCACCTAGG
-5' (序列号 58)

序列号 58 的序列是与序列号 57 的序列互补的序列。

本发明之三提供一种在本发明之一的载体中结合含有如下所述的序
30 列的外来基因而成的重组载体,所述的序列是为了在顺向插入到用第一限

制性酶处理本发明之一的载体后得到的载体时不出现第二限制性酶的切割识别序列，而在逆向插入时出现第二限制性酶的切割识别序列而设计的。优选将外来基因顺向插入到用第一限制性酶处理本发明之一的载体后得到的序列中的重组载体。外来基因可以编码蛋白质和/或蛋白质结构域。

5 本发明之四提供一种含有本发明之三的重组载体的转化体或其后代。

本发明之五提供一种包括如下过程的本发明之一的载体的制造方法：将含有下述两个切割识别序列的双链 DNA 导入到质粒、噬菌体 (phage)、粘粒 (cosmid)、P1 载体、细菌人工染色体载体或酵母人工染色体载体的多克隆位点中，所述的两个切割识别序列是被第一限制性酶识别的碱基序
10 列不同的两个切割识别序列，其中一个切割识别序列含有由不同于第一限制性酶的第二限制性酶所识别的切割识别序列的一部分，而另一个切割识别序列不含由上述第二限制性酶识别的切割识别序列的一部分。

本发明之六提供一种含有下述两个切割识别序列的双链 DNA，所述的两个切割识别序列是被第一限制性酶识别的碱基序列不同的两个切割
15 识别序列，其中一个切割识别序列含有由与第一限制性酶不同的第二限制性酶识别的切割识别序列的一部分，而另一个切割识别序列不含由上述第二限制性酶识别的切割识别序列的一部分。

作为本发明之六的双链 DNA，可以例示为用下述序列 (IX) 和 (X) 表示的双链 DNA。

20 5' -GACX^{1a}X^{2a}TX^{3a}X^{4a}GTC-N¹-GACX^{5a}X^{6a}AX^{7a}X^{8a}GTC-3' (IX)

3' -CTGX^{1b}X^{2b}AX^{3b}X^{4b}CAG-N²-CTGX^{5b}X^{6b}TX^{7b}X^{8b}CAG-5' (X)

(在序列 (IX) 和 (X) 中，G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧啶、T 为胸腺嘧啶，

X^{1a}~X^{8a} 和 X^{1b}~X^{8b} 分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧
25 啶的任意碱基，X^{1a} 与 X^{1b} 为彼此可以形成配对的碱基，X^{2a} 与 X^{2b} 为彼此可以形成配对的碱基，X^{3a} 与 X^{3b} 为彼此可以形成配对的碱基，X^{4a} 与 X^{4b} 为彼此可以形成配对的碱基，X^{5a} 与 X^{5b} 为彼此可以形成配对的碱基，X^{6a} 与 X^{6b} 为彼此可以形成配对的碱基，X^{7a} 与 X^{7b} 为彼此可以形成配对的碱基，X^{8a} 与 X^{8b} 为彼此可以形成配对的碱基，

30 序列 CX^{1a}X^{2a}T 为含有 N c o I、D s a I、E c o T 1 4 I、N d

e I、S f e I、A v a I、S m l I、X h o I、P s t I、E c o T
2 2 I、P s h B I、A c c I、A p a L I、P s i I、S c a I、S
p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B s p T 1 0 7 I、B
s p 1 4 0 7 I、T a t I、H g i C I、K p n I、S p e I、B l n
5 I、N h e I、X b a I、S s p I、E c o R V以及它们的同裂酶的切
割识别序列的5'侧的部分序列的碱基序列，

序列CX^{8b}X^{7b}T为不含在序列CX^{1a}X^{2a}T中含有的N c o I、D s a I、
E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、
P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s
10 i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B
s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、
S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V以及
它们的同裂酶的切割识别序列的部分序列的碱基序列，

5'—和3'—表示双链DNA的各自的末端，

15 N¹可以不存在，但当存在时，为1~1000碱基的任意核苷酸序列，

N²可以不存在，但当存在时，为与N¹的核苷酸序列互补的核苷酸序
列。)

另外，作为本发明之六的双链DNA，可以例示为用下述序列(XI)
和(XII)表示的双链DNA。

20 5' -CCAY^{1a}Y^{2a}Y^{3a}Y^{4a}TY^{5a}Y^{6a}Y^{7a}Y^{8a}TGG-N¹-CCAY^{9a}Y^{10a}Y^{11a}Y^{12a}AY^{13a}Y^{14a}Y^{15a}Y^{16a}
TGG-3' (XI)

3' -GGTY^{1b}Y^{2b}Y^{3b}Y^{4b}AY^{5b}Y^{6b}Y^{7b}Y^{8b}ACC-N²-GGTY^{9b}Y^{10b}Y^{11b}Y^{12b}TY^{13b}Y^{14b}Y^{15b}Y^{16b}
ACC-5' (XII)

(在序列(XI)和(XII)中，G为鸟嘌呤、A为腺嘌呤、C为胞嘧
25 啶、T为胸腺嘧啶，

Y^{1a}~Y^{16a}和Y^{1b}~Y^{16b}分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺
嘧啶的任意碱基，Y^{1a}与Y^{1b}为彼此可以形成配对的碱基，Y^{2a}与Y^{2b}为彼
此可以形成配对的碱基，Y^{3a}与Y^{3b}为彼此可以形成配对的碱基，Y^{4a}与
Y^{4b}为彼此可以形成配对的碱基，Y^{5a}与Y^{5b}为彼此可以形成配对的碱基，

Y^{6a}与Y^{6b}为彼此可以形成配对的碱基, Y^{7a}与Y^{7b}为彼此可以形成配对的碱基, Y^{8a}与Y^{8b}为彼此可以形成配对的碱基, Y^{9a}与Y^{9b}为彼此可以形成配对的碱基, Y^{10a}与Y^{10b}为彼此可以形成配对的碱基, Y^{11a}与Y^{11b}为彼此可以形成配对的碱基, Y^{12a}与Y^{12b}为彼此可以形成配对的碱基, Y^{13a}与Y^{13b}为彼此可以形成配对的碱基, Y^{14a}与Y^{14b}为彼此可以形成配对的碱基, Y^{15a}与Y^{15b}为彼此可以形成配对的碱基, Y^{16a}与Y^{16b}为彼此可以形成配对的碱基,

序列 Y^{1a}Y^{2a}Y^{3a}Y^{4a}T 为含有 N c o I、D s a I、E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V、A f l I I I、B s p L U 1 1 I、N s p I、S p h I、B s p H I、B s a A I、P m a C I、S n a B I、N s p B I I、P v u I I、B a n I I、H g i A I、S a c I 以及它们的同裂酶的切割识别序列的 5'侧的部分序列的碱基序列,

序列 Y^{16b}Y^{15b}Y^{14b}Y^{13b}T 为不含在序列 Y^{1a}Y^{2a}Y^{3a}Y^{4a}T 中含有的 N c o I、D s a I、E c o T 1 4 I、N d e I、S f e I、A v a I、S m l I、X h o I、P s t I、E c o T 2 2 I、P s h B I、A c c I、A p a L I、P s i I、S c a I、S p l I、B s p 1 1 0 7 I、B s p 1 2 8 6 I、B s p T 1 0 7 I、B s p 1 4 0 7 I、T a t I、H g i C I、K p n I、S p e I、B l n I、N h e I、X b a I、S s p I、E c o R V、A f l I I I、B s p L U 1 1 I、N s p I、S p h I、B s p H I、B s a A I、P m a C I、S n a B I、N s p B I I、P v u I I、B a n I I、H g i A I、S a c I 以及它们的同裂酶的切割识别序列的部分序列的碱基序列,

5'—和 3'—表示双链 DNA 的各自的末端,

N¹可以不存在, 但当存在时, 为 1~1000 碱基的任意核苷酸序列,

N²可以不存在, 但当存在时, 为与 N¹的核苷酸序列互补的核苷酸序

列。)

进而,作为本发明之六的双链DNA,可以例示为用下述序列(XIII)和(XIV)表示的双链DNA。

5' -GACX^{1a}X^{2a}TX^{3a}X^{4a}GTC-N¹-CCAY^{9a}Y^{10a}Y^{11a}Y^{12a}AY^{13a}Y^{14a}Y^{15a}Y^{16a}TGG-3' (XIII)

5 3' -CTGX^{1b}X^{2b}AX^{3b}X^{4b}CAG-N²-GGTY^{9b}Y^{10b}Y^{11b}Y^{12b}TY^{13b}Y^{14b}Y^{15b}Y^{16b}ACC-5' (XIV)

(在序列(XIII)和(XIV)中,G为鸟嘌呤、A为腺嘌呤、C为胞嘧啶、T为胸腺嘧啶,

X^{1a}~X^{4a}和X^{1b}~X^{4b}分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基,X^{1a}与X^{1b}为彼此可以形成配对的碱基,X^{2a}与X^{2b}为彼此
10 可以形成配对的碱基,X^{3a}与X^{3b}为彼此可以形成配对的碱基,X^{4a}与X^{4b}为彼此可以形成配对的碱基,

Y^{9a}~Y^{16a}和Y^{9b}~Y^{16b}分别独立地为鸟嘌呤、腺嘌呤、胞嘧啶或胸腺嘧啶的任意碱基,Y^{9a}与Y^{9b}为彼此可以形成配对的碱基,Y^{10a}与Y^{10b}为彼此可以形成配对的碱基,Y^{11a}与Y^{11b}为彼此可以形成配对的碱基,Y^{12a}
15 与Y^{12b}为彼此可以形成配对的碱基,Y^{13a}与Y^{13b}为彼此可以形成配对的碱基,Y^{14a}与Y^{14b}为彼此可以形成配对的碱基,Y^{15a}与Y^{15b}为彼此可以形成配对的碱基,Y^{16a}与Y^{16b}为彼此可以形成配对的碱基,

序列CX^{1a}X^{2a}T为含有NcoI、DsaI、EcoT14I、NdeI、SfeI、AvaI、SmlI、XhoI、PstI、EcoT
20 22I、PshBI、AccI、ApaLI、PsiI、ScaI、SplI、Bsp1107I、Bsp1286I、BspT107I、Bsp1407I、TatI、HgiCI、KpnI、SpeI、BlnI、NheI、XbaI、SspI、EcoRV以及它们的同裂酶的切割识别序列的5'侧的部分序列的碱基序列,

25 序列CY^{14b}Y^{13b}T为不含在序列CX^{1a}X^{2a}T中含有的NcoI、DsaI、EcoT14I、NdeI、SfeI、AvaI、SmlI、XhoI、PstI、EcoT22I、PshBI、AccI、ApaLI、PsiI、ScaI、SplI、Bsp1107I、Bsp1286I、BspT107I、Bsp1407I、TatI、HgiCI、KpnI、
30 I、SpeI、BlnI、NheI、XbaI、SspI、EcoRV

以及它们的同裂酶的切割识别序列的部分序列的碱基序列，

5'—和 3'—表示双链 DNA 的各自的末端。

N¹ 可以不存在，但当存在时，为 1~1000 碱基的任意核苷酸序列，

N² 可以不存在，但当存在时，为与 N¹ 的核苷酸序列互补的核苷酸序列。)

本发明之七提供一种使用本发明之一的载体来制造外来基因编码的蛋白质和/或蛋白质结构域的方法。

在本发明之七的一种方式中，制作将作为对蛋白质和/或蛋白质结构域进行编码且含有下述的序列的外来基因结合在本发明之一的载体中而成的重组载体，用第二限制性酶对该重组载体进行处理之后，导入到宿主中，形成宿主的单菌落，通过培养单菌落来源的宿主，使外来基因所编码的蛋白质和/或蛋白质结构域表达，其中所述的序列被设计成当顺向插入到已用第一限制性酶处理本发明之一的载体得到的载体时不出现第二限制性酶的切割识别位点，当逆向插入时出现第二限制性酶的切割识别序列。

在本发明之七的其它方式中，可以使用含有如下所述的序列并能对编码蛋白质和/或蛋白质结构域的外来基因的 cDNA 进行扩增的引物，以 cDNA 作为模板，对 cDNA 中的位置和/或长度以及/或序列的一部分不同的相当于 cDNA 的部分序列的 DNA 的混合物进行扩增，制作将各混合物结合到用第一限制性酶处理过的本发明之一的载体中的重组载体，用第二限制性酶处理过重组载体之后，导入到宿主中，形成宿主的单菌落，通过培养单菌落来源的宿主，使外来基因编码的蛋白质和/或蛋白质结构域表达，所述的序列被设计为当顺向插入到已用第一限制性酶处理本发明之一的载体得到的载体时不出现第二限制性酶的切割识别序列，当逆向插入时出现第二限制性酶的切割识别序列。

本发明之八提供一种使用本发明之一的载体，以使外来基因被逆向插入到载体中而成的重组载体开环、将外来基因被顺向插入到载体中而成的重组载体闭环的状态进行保留的方法。

在本发明之八的一种方式中，通过用第二限制性酶处理本发明之三的重组载体，能够以使外来基因被逆向插入到载体中而成的重组载体开环、

将外来基因被顺向插入到载体中而成的重组载体闭环的状态进行保留。

本发明之九提供一种使用本发明之一的载体,对外来基因所编码的蛋白质和/或蛋白质结构域的表达量和/或溶解度进行试验的方法。

在本发明之九的一种方式中,可以制作将对蛋白质和/或蛋白质结构域进行编码且含有下述序列的外来基因结合到本发明之一的载体中而成的重组载体,在用第二限制性酶处理过该重组载体之后,导入到宿主中,培养宿主和/或其后代,由此使外来基因所编码的蛋白质和/或蛋白质结构域表达,可以检查已表达的蛋白质和/或蛋白质结构域的量 and/或溶解度,所述的序列被设计成当顺向插入到用第一限制性酶处理过的本发明之一的载体时不出现第二限制性酶的切割识别序列,当逆向插入时出现第二限制性酶的切割识别序列。

将序列(I)显示于序列表的序列号1中。

将序列(II)显示于序列表的序列号2中。

15 将序列(III)显示于序列表的序列号3中。

将序列(IV)显示于序列表的序列号4中。

将序列(V)显示于序列表的序列号5中。

将序列(VI)显示于序列表的序列号6中。

将序列(VII)显示于序列表的序列号7中。

20 将序列(VIII)显示于序列表的序列号8中。

将序列(IX)的一个例子(N^1 为一个碱基时)显示于序列表的序列号9中。

将序列(X)的一个例子(N^2 为一个碱基时)显示于序列表的序列号10中。

25 将序列(XI)的一个例子(N^1 为一个碱基时)显示于序列表的序列号11中。

将序列(XII)的一个例子(N^2 为一个碱基时)显示于序列表的序列号12中。

30 将序列(XIII)的一个例子(N^3 为一个碱基时)显示于序列表的序列号13中。

将序列 (XIV) 的一个例子 (N^4 为一个碱基时) 显示于序列表的序列号 14 中。

将序列 (XV) 的一个例子 (N^3 为一个碱基时) 显示于序列表的序列号 15 中。

5 将序列 (XVI) 的一个例子 (N^4 为一个碱基时) 显示于序列表的序列号 16 中。

将序列 (XVII) 的一个例子 (N^3 为一个碱基时) 显示于序列表的序列号 17 中。

10 将序列 (XVIII) 的一个例子 (N^4 为一个碱基时) 显示于序列表的序列号 18 中。

将序列 (XIX) 的一个例子 (N^3 为一个碱基时) 显示于序列表的序列号 19 中。

将序列 (XX) 的一个例子 (N^4 为一个碱基时) 显示于序列表的序列号 20 中。

15 在本说明书中,“蛋白质结构域”是蛋白质的部分序列,是指只有该部分具有某种生理活性、生化学功能、特征性的物理化学性质或者立体结构中的任意性质的多肽。

“(外来基因)被顺向插入”是指该基因中含有的 ORF 相对质粒的翻译方向以正向被插入。

20 “(外来基因)被逆向插入”是指该基因中含有的 ORF 相对质粒的翻译方向以与正向相反的方向被插入。

“切割识别序列”是指各限制性酶中固有的、由该限制性酶识别并将其切割的双链 DNA 的碱基序列。

“限制性酶位点”是指特定的限制性酶的“切割识别序列”。

25 “多克隆位点”是指被设计成多个限制性酶位点连续出现的载体中的碱基序列。

“转化体”是指从细胞向外部导入基因 (DNA)、结果细胞获得新的遗传特征、进而该遗传特征也被后代细胞接受继承。

“宿主”是指在重组 DNA 实验中,导入 DNA 的重组分子的活细胞。

30 “载体”是指在重组 DNA 实验中,向宿主运输异种 DNA 的 DNA。

“单菌落来源的宿主（或者转化体）”是指充分稀释活细胞后接种于培养基中，进行培养形成菌落之后，分离单一菌落进而培养的宿主（或者转化体）。

通过本发明，提供一种能够通过简单的操作在短时间内使蛋白质和/或蛋白质结构域表达的载体。

另外，通过本发明，还提供一种能够通过简单的操作在短时间内使蛋白质和/或蛋白质结构域表达的方法。

进而，通过本发明，还提供一种能够通过简单的操作在短时间内对蛋白质和/或蛋白质结构域的表达量和溶解度进行试验的方法。

本说明书包括作为本申请的优先权基础的日本专利申请、特愿 2003-308773 号的说明书和/或附图中记载的内容。

附图说明

图 1 是表示 dGST-T 载体（pGEX-4T3-PRESAT）的构建。

图 2 是表示 ORF 方向性选择的原理图。图 2 表示第二限制性酶为 Nco I 的情况，反向 PCR 引物（rear PCR primer）被设计成具有 5'-GG。当将利用该引物扩增的插入序列（insert）以反方向插入到质粒中时，在 T/A 克隆位点出现 Nco I 部位。当第二限制性酶为 Nde I 时，反向 PCR 引物被设计成具有 5'-ATG 来代替 5'-GC。

图 3 是表示利用了 dGST-T 载体的可溶性蛋白质结构域的结构域边界的高速确定法的示意图。（i）设计各种 PCR 引物，扩增作为一系列目的的部分 cDNA 片断。推断该部分 cDNA 片断含有蛋白质结构域。（ii）将该片断插入到 pGEX-4T3-PRESAT 中，进行 Nco I 或 Nde I 选择，转换到大肠杆菌，分离大肠杆菌的菌落。（iii）使用 96 孔板，培养各个菌落，使各种片断长度的靶蛋白质和/或蛋白质结构域的 GST 融合蛋白质表达。

（iv）利用 CDNB 显色反应对融合蛋白质的量进行定量，选择高表达 GST 融合蛋白质的克隆，将其用于进一步的分析中。

图 4 是表示用 15% SDS-PAGE 对 GST 融合蛋白质的表达进行分析的结果。M 道（lane）表示分子量标准，a1~a5 道表示酵母 Vps4，b1~b5 道表示人 Vps4，c1~c5 道表示小鼠 Vps4，d1~d5 道表示人 Snx15a，

e1~e5 道表示小鼠 Snx15a。

图 5 是表示用 15% SDS-PAGE 对 GST 融合蛋白质在菌体刚刚破碎
碎裂之后的溶解度进行分析的结果。M 道表示分子量标准, i1~s1 道表示
酵母 Vsp4, i2~s2 道表示人 Vps4, i3~s3 道表示小鼠 Vps4, i4~s4 道表
5 示人 Snx15a, i5~s5 道表示小鼠 Snx15a。另外, 各道的 i 表示不溶性组
分, s 表示可溶性组分。

图 6 是表示利用作为 GST 的酶活性测量法的使用了 1-氯-2, 4-
二硝基苯的检验方法来检测 GST 融合蛋白质在菌体刚刚碎裂之后的溶解
度的结果。图中横轴表示从反应开始起的时间(单位: 分钟), 纵轴表示
10 在 350nm 下的紫外线的吸光度。另外, 图表中△表示酵母 Vps4, ○表示
人 Vps4, ●表示小鼠 Vps4, □表示人 Snx15a, ▲表示小鼠 Snx15a。

图 7 是表示利用了 dGST-T 载体的蛋白质结构域的高速可溶性试验
法的示意图。[]内是指在此记述的操作中所必要的大概时间。(i) 设计
特定的 PCR 引物, 对作为目的的部分 cDNA 片断进行扩增。该部分 cDNA
15 片断可以是推断为含整个蛋白质或蛋白质结构域的部分序列的任意片
断。(ii) 将该片断插入到 pGEX-4T3-PRESAT 中, 进行连接(ligation)
反应。(iii) 转换到大肠杆菌 DH5a。(iv) 进行液体培养。(v) 分离纯化
质粒。(vi) 用 Nco I 或 Nde I 切割而进行 ORF 选择。(vii) 转换到大肠杆
菌 BL21, (viii) 使大肠杆菌的菌落形成。(ix) 对 IPTG 溶液进行喷雾, 诱
20 导蛋白质的表达。(x) 悬浮琼脂培养基上的大肠杆菌并收集。(xi) 使
菌体碎裂, 得到全菌体、可溶性组分、不溶性组分。(xii) 进行 SDS
-PAGE 电泳, 通过凝胶上的带(band)的浓度判断在可溶性组分中含有
GST 融合蛋白质的量。(i)~(xii) 的全部工序需要的时间总计大
致为 48 小时。

25 图 8 是表示用 TA-克隆法将编码 mVps4b 和 mSNX15a 的结构域的
cDNA 插入到 pGST-4T3-PRESAT 中, 经过 Nco I ORF 选择法使得在琼
脂培养基上形成得菌落不进行单菌落化, 而是在琼脂培养基上进行表达诱
导, 此后进行可溶性试验的结果。M 道表示分子量标志(marker); t1、
s1、i1 道表示使未进行 GST-小鼠 Vps4 融合蛋白质的表达诱导的样品泳
30 动的结果; t2、s2、i2 道表示使利用 IPTG 喷雾诱导该蛋白质表达的样品

泳动的结果；t3、s3、i3 道表示使未进行 GST—小鼠 SNX15 融合蛋白质的表达诱导的样品泳动的结果；t4、s4、i4 道表示使利用 IPTG 喷雾诱导该蛋白质表达的样品泳动后的结果，t、i、s 分别表示全菌体、不溶性组分、可溶性组分。另外，图中箭头表示已确认表达的融合蛋白质的带，△表示只有 GST 的带。

具体实施方式

下面对本发明的优选实施方式进行详细说明。

1. 载体的制备

首先，准备载体。例如购入市售的载体，在该载体具有由第一和第二限制性酶识别的切割识别序列时，可以在该切割识别序列中导入突变。如果该载体为融合蛋白质表达载体，从易于检测/测量目的蛋白质和/或蛋白质结构域的表达量、溶解性的观点出发，从易于纯化目的蛋白质和/或蛋白质结构域的观点出发，从当用于进行目的蛋白质的功能分析、相互作用分析的生化实验时有利的观点出发，这是方便的。在后述的实施例 1 中，使用已对用序列号 21 的序列表示的 DNA 进行 5'磷酸化的序列，利用定位突变法（site-directed mutagenesis），向 pGEX—4T3 载体中导入突变。这样，pGEX—4T3 在原来的序列中具有 Ahd I 位点不再被切割。

接着，准备用于向上述载体中导入由第一限制性酶识别的碱基序列不同的两个切割识别序列（其中只有一个切割识别序列含有由与第一限制性酶不同的第二限制性酶所识别的切割识别序列的一部分）的连接序列（linker）。该连接序列在两个末端具有用于向载体的多克隆位点中导入的限制性酶部位（两处），被设计成具有由第一限制性酶识别的碱基序列不同的两个切割识别序列（其中只有一个切割识别序列含有由与第一限制性酶不同的第二限制性酶识别的切割识别序列的一部分）。在后述的实施例 1 中，通过将 pET32a 载体的衍生物 pET32aPACAP 作为模板，以序列号 22 和 24 的序列表示的引物对进行 PCR 反应，在两个末端具有 BamH I 部位，得到分别在该两个 BamH I 部位上结合 Ahd I 部位的 PCR 产物。将该 PCR 产物亚克隆到 TA 克隆载体 pGEM—T 中，在已确认其序列之后，制备纯化质粒，用限制性酶 BamH I 切割已纯化的质粒，纯化约 500base 的

片断 (Ahd I 连接序列)。

用限制性酶 (在后述的实施例 1 中为 BamH I) 使上述载体 (在后述的实施例 1 中, 已导入突变的 pGEX-4T3 载体) 开环, 去磷酸化处理之后, 连接 (ligation) 上述的连接序列 (在后述的实施例 1 中为 Ahd I 连接序列), 转换到大肠杆菌。然后, 从多个单菌落来源的转化体制备纯化质粒, 确认其碱基序列, 选择以正向导入连接序列的载体, 用第一限制性酶 (在后述的实施例 1 中为 Ahd I) 对其进行处理, 此时得到开环状载体 (在后述的实施例 1 中为 pGEX-4T3-PRESAT)。

2. 外来基因的制备

10 制备外来基因。外来基因含有如下所述的序列, 即所述的序列被设计成当向在 1 中准备的载体中顺向插入时, 不出现第二限制性酶的切割识别序列, 但当逆向插入时, 出现第二限制性酶的切割识别序列。另外, 外来基因也可以编码蛋白质和/或蛋白质结构域。

外来基因例如可以用 PCR 法制备。作为模板, 可以例示为含有目的基因的 cDNA、质粒、合成 DNA、RNA 等。作为引物, 可以在需要扩增的目的基因的两个末端的 DNA 序列使用互补的两种寡核苷酸, 但需要预先进行正确设计, 并使得一方的引物在插入到载体之后翻译的读码框不错位, 另外另一方的引物在其 5' 末端附加 1~10 个残基的目的基因中不含的序列。

20 PCR 反应是通过使 DNA 双链的解离、与寡核苷酸的退火、以及通过 DNA 聚合酶的互补链的合成三个反应重复进行 20~40 次来进行的。在第一次反应中, 可以通过使用具有 DNA 的逆转录活性的酶, 进行从 RNA 开始的 DNA 扩增。

引物的碱基序列被设计成为: 当被顺向插入到在 1 中准备的载体中时, 不出现第二限制性酶的切割识别序列, 但当被逆向插入时出现第二限制性酶的切割识别序列。在后述的实施例 2 中, 正向引物 (front primer) (与该 cDNA 的 ORF 的 N 末端相当的一侧的引物) 被设计成以最初的氨基酸的密码子开始并与前方的 GST 为读码框一致, 5' 末端的碱基不以 5'-GG 开始。这样, 用这些引物扩增出的 PCR 产物就会只在与 pGEX-4T3 25 -PRESAT 逆向连接时形成 Nco I 部位。该新产生的 Nco I 部位被用在接

下来的选择工序中。

此外，下面，对在设计用于导入到 pGEX-4T3-PRESAT 中并产生重组蛋白质的、用于扩增 DNA 的 PCR 引物的情况下的注意点进行说明。

5 (1) 当第一限制性酶为 Ahd I、第二限制性酶为 Nco I 时 (Nco I 选择)

为了设计 ORF 的 N 末端侧的引物，可以按照以下的四个规则进行。

(i a) 引物序列的第一个碱基号与氨基酸密码子的第一个字对应并与前面的 GST 读框一致。

10 (ii a) 5'末端的碱基设为 G 或 A。(必要时，插入编码少数氨基酸的 DNA 序列)。

(iiia) 5'末端的序列不为 GG。

(iva) 在 PCR 扩增片断内不含有 Nco I 限制性酶位点。

ORF 的 C 末端侧的引物可以按照下面的两个规则设计。

(v a) 5'末端的碱基设为 GG。

15 (via) 当目的蛋白质结构域不含有终止密码子时，5'末端的序列以 GG-Xn-TTA (Xn 可以为任意的 n 个 (n 为 0~30 的任意整数) 碱基，例如为 A。)(序列号 53 中显示 n 为 5 时的序列) 开始。

(2) 当第一限制性酶为 Ahd I、第二限制性酶为 Nde I 时 (Nde I 选择)

20 为了设计 ORF 的 N 末端侧的引物，可以按照以下四个规则。

(i b) 引物序列的第一个碱基与氨基酸密码子的第一个字对应并与前面的 GST 读框一致。

(ii b) 5'末端的碱基设为 G 或 A。(必要时，插入编码少数氨基酸的 DNA 序列)。

25 (iiib) 5'末端的序列不为 ATG。

(ivb) 在 PCR 扩增片断内不含有 Nde I 限制性酶位点。

ORF 的 C 末端侧的引物可以按照下面的两个规则设计。

(v b) 5'末端的碱基设为 ATG。

30 (vib) 当目的蛋白质结构域不含有终止密码子时，5'末端的碱基以 ATG-Xn-TTA (Xn 可以为任意的 n 个 (n 为 0~30 的任意整数) 碱基，

例如为 A。) (序列号 54 中显示 n 为 4 时的序列) 开始。

3. 克隆

使在 2 中准备的外来基因与在 1 中准备的载体 (在后述的实施例 2 中, 为 pGEX-4T3-PRESAT 载体) 发生连接反应, 转换到大肠杆菌。

- 5 不将已转化的大肠杆菌接种于琼脂培养基中, 另外在不进行单菌落化的情况下, 用液体培养基进行培养之后, 从中纯化质粒混合物。当用第二限制性酶 (在后述的实施例 2 中为 Nco I) 处理该质粒混合物时, 外来基因被逆向插入到载体中的质粒在新出现的第二限制性酶部位被切割而开环。另一方面, 外来基因被顺向插入到载体的质粒仍然保留闭环状态。因而, 当
- 10 将用第二限制性酶处理过的质粒混合物转化成表达用大肠杆菌宿主并播种于琼脂培养基中时, 只有含外来基因被顺向插入到载体的质粒的大肠杆菌形成菌落。从已使该单菌落形成的琼脂培养基的各菌落采集大肠杆菌进行培养, 使融合蛋白质表达。收集表达结束后的大肠杆菌, 用超声波等破碎菌体之后, 作为全菌体组分。另外离心分离全菌体组分, 将上清作为可
- 15 溶性组分。将沉淀物质完全溶解于容量与可溶性组分相同的含 SDS 缓冲液中, 作为不溶性组分。利用 SDS-PAGE 只使全菌体组分泳动, 或者分别使全菌体组分、可溶性组分、不溶性组分泳动, 在凝胶上确认目的融合蛋白质的带。然后, 从该融合蛋白质的凝胶上的带的浓度求得可溶性组分和不溶性组分的比率, 作为目的融合蛋白质的溶解度的指标 (可溶性试
- 20 验)。必要时, 利用光密度分析法 (densitometry) 等, 对凝胶上的蛋白质带进行定量, 再将其数值化即可。

- 利用 GST 的酶活性的可溶性试验, 是按照 Amersham Pharmacia 公司的 pGEX 系列载体操作说明书记载的方法进行。即, 将用适当的缓冲液稀释的含有目的融合蛋白质的大肠杆菌的可溶性组分加入到紫外分光光度
- 25 计的比色杯 (cuvette) 中, 加入 1-氯-2,4-二硝基苯和还原型谷胱甘肽, 充分搅拌之后, 连续 5~10 分钟测量 340nm 的紫外吸光度的经时变化, 将其斜率作为目的融合蛋白质的量的指标。在并行处理多个样品时, 如果利用微板读出器 (マイクロプレートリーダー) 则会更方便。

- 在进行上述操作之后, 选择目的融合蛋白质的表达量/溶解度好的克
- 30 隆, 纯化质粒, 确定碱基序列。将含有目的外来基因的质粒用作蛋白质和

/或蛋白质结构域表达载体，将其用于以后的蛋白质和/或蛋白质结构域的大量制备实验中。通过该方法，与以往的方法相比，可以使制作蛋白质和/或蛋白质结构域表达载体的时间缩短约三成。

4. 蛋白质和/蛋白质结构域的大量表达

5 将在 3 中使用的蛋白质和/或蛋白质结构域表达载体转化到表达用大肠杆菌宿主中，培养大肠杆菌，使蛋白质和/或蛋白质结构域表达。当蛋白质和/或蛋白质结构域表达载体为融合蛋白质表达载体时，从易于检测/测量目的蛋白质和/或蛋白质结构域的表达量/溶解性（例如，用荧光进行的检测、将 GST 的酶活性作为指标的比色定量等）的观点出发，从易于
10 纯化目的蛋白质和/或蛋白质结构域（例如利用 GSH 固定化珠子的一個阶段的纯化成为可能）的观点出发，从当进行用于目的蛋白质的功能分析、相互作用分析的生化实验（例如利用 BIACORE 的试管内结合试验、利用 pull down 检测的试管内结合试验等）时有利的观点出发，这是方便的。

已表达的蛋白质和/或蛋白质结构域的分离和纯化可以利用公知的方法进行。作为公知的分离、纯化方法，可以使用盐析或溶剂沉淀法等利用溶解度的方法，透析法、超滤法、凝胶过滤法以及 SDS-聚丙烯酰胺凝胶电泳法等利用分子量的差的方法，离子交换层析法（chromatography）等
15 利用电荷的差的方法，亲和层析（affinity chromatography）等利用特异亲和性的方法，逆相高速液体层析或疏水层析等利用疏水性的差的方法，等
20 电点电泳法等利用等电点的差的方法等。

以上，是针对本发明中将大肠杆菌作为宿主使用时的情况进行了说明，但本发明中使用的宿主并不限于大肠杆菌，作为宿主，可以例示为细菌细胞（例如埃希氏（*Escherichia*）菌属、芽孢杆菌属、枯草杆菌等）、真菌细胞（例如，出芽酵母、分裂酵母、曲霉属等）、昆虫细胞（例如，S2
25 细胞、Sf 细胞等）、动物细胞（例如，CHO 细胞、COS 细胞、HeLa 细胞、C127 细胞、3T3 细胞、BHK 细胞、HEK 293 细胞等）、植物细胞等。

另外，本发明中使用的第一限制性酶、第二限制性酶、由第一限制性酶识别的碱基序列不同的两个切割识别序列（含有被第二限制性酶识别的切割识别序列的一部分的序列和不含有上述的序列）、用于制作外来基因的
30 引物对均不被上述所限定，可以利用下面表中所归纳的酶或序列等。表

中，G 为鸟嘌呤、A 为腺嘌呤、C 为胞嘧啶、T 为胸腺嘧啶、X 为任意碱基、X_n 为 n 个（n 为 1~30 的任意整数）任意碱基。表 2 和表 3 接表 1，表 5 和表 6 接表 4，表 8~11 接表 7。

表 1

第一限制酶	第一限制酶的切割识别序列① (含有第二限制酶的切割识别序列的一部分 (下划线部分))	第一限制酶的切割识别序列② (不含第二限制酶的切割识别序列的一部分)	第二限制酶	第二限制酶的切割识别序列	导入到载体中的外来基因的正向引物的5'末端侧的序列(大A为在PCR反应中附加的碱基)	用于制作外来基因的反向引物的5'末端侧附加的序列(不附加终止密码子的情况)	在用于制作外来基因的反向引物的5'末端附加的序列(也包含终止密码子附加的情况)
AndI	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	NcoI DsaI Eco114I	5'-CCAATGG-3' 3'-GGTACC-5'	5'-xxxx-// -xgcga-3' 3'-Axxxx-// -xgcga-5' (X为任意碱基)	不含有5' -GG	5'-GG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	NdeI	5'-CAATATG-3' 3'-GTATAC-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -ATG	5'-ATG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	SfeI	5'-CTATAG-3' 3'-GATATC-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -ATAG	5'-ATAG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	SfeI	5'-CTATAG-3' 3'-GATATC-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -AG	5'-AG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	MunI	5'-CAATTG-3' 3'-GTTAAC-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -TG	5'-TG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	McrI PvuI	5'-CGATGG-3' 3'-GCTAGC-5'	5'-xxxx-// -xgcga-3' 3'-Axxxx-// -xgcga-5' (X为任意碱基)	不含有5' -CG	5'-CG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	AvaI SmlI XhoI	5'-CTCGAG-3' 3'-GAGCTC-5'	5'-xxxx-// -xgcga-3' 3'-Axxxx-// -xgcga-5' (X为任意碱基)	不含有5' -CGAG	5'-CGAG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	PstI SfeI	5'-CTGCAG-3' 3'-GAGCTC-5'	5'-xxxx-// -xgcga-3' 3'-Axxxx-// -xgcga-5' (X为任意碱基)	不含有5' -GCAG	5'-GCAG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	AflII SmlI	5'-CTTAAG-3' 3'-GAATTC-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -TAAG	5'-TAAG (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	ClaI	5'-ATCGAT-3' 3'-TAGCTA-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -CGAT	5'-CGAT (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	Eco122I	5'-ATGCAT-3' 3'-TAGCTA-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -GCAT	5'-GCAT (Xn为n个任意碱基)
	5'-GACCAATGGTC-3' 3'-CTGGTAACGAG-5'	5'-GACAAATGTC-3' 3'-CTGTTTTACAG-5'	PshI	5'-ATTAAAT-3' 3'-TAATTA-5'	5'-xxxx-// -xgcata-3' 3'-Axxxx-// -xgcata-5' (X为任意碱基)	不含有5' -TAAT	5'-TAAT (Xn为n个任意碱基)

表 2

5'-GACATTTGGTC-3' 3'-CTGTAACACG-5'	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Psh I	5'-ATTAA-3' 3'-TAATTA-5'	5'-XXXX-// -XXATA-3' 3'-AXXXX-// -XXTAA-5' (X 为任意碱基)	不含有 5' -AAT	5'-AAT	5'-AATXnTAA (Xn 为 n 个任意碱基)
5'-GACXGTTGGTC-3' 3'-CTGCAACACG-5' (X 为任意碱基)	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Acc I Bst II 107 I	5'-GTATAC-3' 3'-CATATG-5'	5'-XXXX-// -XGCTAA-3' 3'-AXXXX-// -XGCTAA-5' (X 为任意碱基)	不含有 5' -ATAC	5'-ATAC	5'-ATACXnTAA (Xn 为 n 个任意碱基)
5'-GACXGTTGGTC-3' 3'-CTGCAACACG-5' (X 为任意碱基)	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Acc I Hinc II Sal I	5'-GTGAC-3' 3'-CAGCTG-5'	5'-XXXX-// -XGCTGA-3' 3'-AXXXX-// -XGCTGA-5' (X 为任意碱基)	不含有 5' -CGAC	5'-CGAC	5'-CGACXnTAA (Xn 为 n 个任意碱基)
5'-GACXGTTGGTC-3' 3'-CTGCAACACG-5' (X 为任意碱基)	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Apa I Bsp I 286 I Hgi A I	5'-GTGAC-3' 3'-CAGCTG-5'	5'-XXXX-// -XGCTGA-3' 3'-AXXXX-// -XGCTGA-5' (X 为任意碱基)	不含有 5' -GCAC	5'-GCAC	5'-GCACXnTAA (Xn 为 n 个任意碱基)
5'-GACXGTTGGTC-3' 3'-CTGCAACACG-5' (X 为任意碱基)	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Hinc II Hpa I	5'-GTAAAC-3' 3'-CAATTG-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -TAAC	5'-TAAC	5'-TAACXnTAA (Xn 为 n 个任意碱基)
5'-GACXGTTGGTC-3' 3'-CTGCAACACG-5' (X 为任意碱基)	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Psi I	5'-TTATAA-3' 3'-AATATT-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -ATAA	5'-ATAA	5'-ATAAXnTAA (Xn 为 n 个任意碱基)
5'-GACXGTTGGTC-3' 3'-CTGCAACACG-5' (X 为任意碱基)	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Nsp V Bsp I 104 I	5'-TTGAA-3' 3'-AAGCTT-5'	5'-XXXX-// -XGCTGA-3' 3'-AXXXX-// -XGCTGA-5' (X 为任意碱基)	不含有 5' -CGAA	5'-CGAA	5'-CGAAXnTAA (Xn 为 n 个任意碱基)
5'-GACXGTTGGTC-3' 3'-CTGCAACACG-5' (X 为任意碱基)	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Dra I	5'-TTTAA-3' 3'-AATTT-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -TAAA	5'-TAAA	5'-TAAAXnTAA (Xn 为 n 个任意碱基)
5'-GACGTTGGTC-3' 3'-CTGCAACACG-5'	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Sca I Tat I	5'-AGTACT-3' 3'-TCATGA-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -ACT	5'-ACT	5'-ACTXnTAA (Xn 为 n 个任意碱基)
5'-GACGTTGGTC-3' 3'-CTGCAACACG-5'	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Spl I	5'-CGTACG-3' 3'-GCATGC-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -ACG	5'-ACG	5'-ACGXnTAA (Xn 为 n 个任意碱基)
5'-GACGTTGGTC-3' 3'-CTGCAACACG-5'	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Bsp I 107 I Hgi C I Kpn I	5'-GGTACG-3' 3'-CCATGG-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -ACC	5'-ACC	5'-ACGXnTAA (Xn 为 n 个任意碱基)
5'-GACGTTGGTC-3' 3'-CTGCAACACG-5'	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Bsp I 1407 I Tat I	5'-TGTACA-3' 3'-ACATGT-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -ACA	5'-ACA	5'-ACAXnTAA (Xn 为 n 个任意碱基)
5'-GACGTTGGTC-3' 3'-CTGCAACACG-5'	5'-GACAAATGTC-3' 3'-CTGTTTAAACG-5'	Spe I	5'-ACTAGT-3' 3'-TGATCA-5'	5'-XXXX-// -XGCTTA-3' 3'-AXXXX-// -XGCTTA-5' (X 为任意碱基)	不含有 5' -AGT	5'-AGT	5'-AGTXnTAA (Xn 为 n 个任意碱基)

表 3

5'-GACCCITGGTC-3' 3'-CTGGGACCAG-5'	5'-GACAAAATGTC-3' 3'-CTGTTTACAG-5'	Bln I Eco114I	5'-CCTAGG-3' 3'-GGATCC-5'	5'-XXXX-// -XXCTA-3' 3'-Axxxx-// -XXGGA-5' (X 为任意碱基)	不含有 5' -AGG	5' -AGG	5'-AGGXnTTA (Xn 为 n 个任意碱基)
5'-GACGCTGGTC-3' 3'-CTGGGACCAG-5'	5'-GACAAAATGTC-3' 3'-CTGTTTACAG-5'	Nhe I	5'-GCTAGC-3' 3'-CGATCG-5'	5'-XXXX-// -XXCTA-3' 3'-Axxxx-// -XXGGA-5' (X 为任意碱基)	不含有 5' -AGC	5' -AGC	5'-AGCXnTTA (Xn 为 n 个任意碱基)
5'-GACCTCTGGTC-3' 3'-CTGAGACCAG-5'	5'-GACAAAATGTC-3' 3'-CTGTTTACAG-5'	Xba I	5'-TCTAGA-3' 3'-AGATCT-5'	5'-XXXX-// -XXCTA-3' 3'-Axxxx-// -XXGGA-5' (X 为任意碱基)	不含有 5' -AGA	5' -AGA	5'-AGAXnTTA (Xn 为 n 个任意碱基)
5'-GACAAITGGTC-3' 3'-CTGTTTACCAG-5'	5'-GACAAAATGTC-3' 3'-CTGTTTACAG-5'	Ssp I	5'-AATATT-3' 3'-TTATTA-5'	5'-XXXX-// -XXAAT-3' 3'-Axxxx-// -XXTTA-5' (X 为任意碱基)	不含有 5' -ATT	5' -ATT	5'-ATTXnTTA (Xn 为 n 个任意碱基)
5'-GACGAIITGGTC-3' 3'-CTGCTAACCCAG-5'	5'-GACAAAATGTC-3' 3'-CTGTTTACAG-5'	EcoRV	5'-GATATC-3' 3'-CTATAG-5'	5'-XXXX-// -XXGAT-3' 3'-Axxxx-// -XXGTA-5' (X 为任意碱基)	不含有 5' -ATC	5' -ATC	5'-ATCXnTTA (Xn 为 n 个任意碱基)

表 4

第一限制酶	第一限制酶的切割识别序列① (含有第二限制酶的切割识别序列的一部分(下划线部分))	第一限制酶的切割识别序列② (不含第二限制酶的切割识别序列的一部分)	第二限制酶	第二限制酶的切割识别序列	导入到载体中的外来基 因的 5' 末端侧和 3' 末端侧的序列 (大 A 为在 PCR 反应中 附加的碱基)	用于制作外来基 正向引物的 5' 末端 序列应该满足的条件	在用于制作外来 基因的反向引物 的 5' 末端附加 的序列 (不附加 终止密码子的情 况)	在用于制作外来基 因的反向引物 的 5' 末端附加的 序列 (也包括终止密码子 附加的情况)
AhdI/XcmI	5'-GACCAITGGTC-3' 3'-CTGGTAACCCAG-5'	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	NcoI DsaI EcoRII	5'-CCATGG-3' 3'-GGTACC-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXXX-5' (X 为任意碱基)	不含有 5' -GG	5'-GG (Xn 为 n 个任意碱基)	
	5'-GACCAITGGTC-3' 3'-CTGGTAACCCAG-5'	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	NdeI	5'-CATATG-3' 3'-GTATAC-5'	5'-XXXX-// -XXCATA-3' 3'-AXXXX-// -XXGTA-5' (X 为任意碱基)	不含有 5' -ATG	5'-ATG (Xn 为 n 个任意碱基)	5'-ATGXnTTA (Xn 为 n 个任意碱基)
	5'-GACXCITGGTC-3' 3'-CTGXGAACCCAG-5' (X 为任意碱基)	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	SfeI	5'-CTATAG-3' 3'-GATATC-5'	5'-XXXX-// -XXCTA-3' 3'-AXXXX-// -XXGTA-5' (X 为任意碱基)	不含有 5' -ATAG	5'-ATAG (Xn 为 n 个任意碱基)	5'-ATAGXnTTA (Xn 为 n 个任意碱基)
	5'-GACITAITGGTC-3' 3'-CTGATAACCCAG-5'	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	SfeI	5'-CTATAG-3' 3'-GATATC-5'	5'-XXXX-// -XXCTA-3' 3'-AXXXX-// -XXGTA-5' (X 为任意碱基)	不含有 5' -AG	5'-AG (Xn 为 n 个任意碱基)	5'-AGXnTTA (Xn 为 n 个任意碱基)
	5'-GACAAITGGTC-3' 3'-CTGTTAACCCAG-5'	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	MunI	5'-CAATTG-3' 3'-GTTAAC-5'	5'-XXXX-// -XXCAA-3' 3'-AXXXX-// -XXGT-5' (X 为任意碱基)	不含有 5' -TG	5'-TG (Xn 为 n 个任意碱基)	5'-TGXnTTA (Xn 为 n 个任意碱基)
	5'-GACGAIITGGTC-3' 3'-CTGCTAACCCAG-5'	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	McrI PvuI	5'-CGATGG-3' 3'-GCTAGC-5'	5'-XXXX-// -XXCDA-3' 3'-AXXXX-// -XXGC-5' (X 为任意碱基)	不含有 5' -CG	5'-CG (Xn 为 n 个任意碱基)	5'-CGXnTTA (Xn 为 n 个任意碱基)
	5'-GACXCITGGTC-3' 3'-CTGXGAACCCAG-5' (X 为任意碱基)	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	AvaI SmlI XhoI	5'-CTCGAG-3' 3'-GAGCTC-5'	5'-XXXX-// -XXCTCA-3' 3'-AXXXX-// -XXGAC-5' (X 为任意碱基)	不含有 5' -CGAG	5'-CGAG (Xn 为 n 个任意碱基)	5'-CGAGXnTTA (Xn 为 n 个任意碱基)
	5'-GACXCITGGTC-3' 3'-CTGXGAACCCAG-5' (X 为任意碱基)	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	PstI SfeI	5'-CTGCAAG-3' 3'-GACGTC-5'	5'-XXXX-// -XXCTCA-3' 3'-AXXXX-// -XXGAC-5' (X 为任意碱基)	不含有 5' -GCAG	5'-GCAG (Xn 为 n 个任意碱基)	5'-GCAGXnTTA (Xn 为 n 个任意碱基)
	5'-GACXCITGGTC-3' 3'-CTGXGAACCCAG-5' (X 为任意碱基)	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	AflII SmlI	5'-CTTAAG-3' 3'-GAATTC-5'	5'-XXXX-// -XXCTTA-3' 3'-AXXXX-// -XXGAAT-5' (X 为任意碱基)	不含有 5' -TAAG	5'-TAAG (Xn 为 n 个任意碱基)	5'-TAAGXnTTA (Xn 为 n 个任意碱基)
	5'-GACXALITGGTC-3' 3'-CTGXIAACCCAG-5' (X 为任意碱基)	5'-CCAAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	ClaI	5'-ATCGAT-3' 3'-TAGCTA-5'	5'-XXXX-// -XXATGA-3' 3'-AXXXX-// -XXTAC-5' (X 为任意碱基)	不含有 5' -CGAT	5'-CGAT (Xn 为 n 个任意碱基)	5'-CGATXnTTA (Xn 为 n 个任意碱基)

表 5

5'-GACXALIGGTC-3' 3'-CTGXTAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	EcoT22I	5'-ATGCAT-3' 3'-TACGTA-5'	5'-XXXX-// -XXATGA-3' 3'-AXXX-// -XTTACG-5' (X为任意碱基)	不含有 5' -GCAT	5'-GCAT	5'-GCATXnTTA (Xn为n个任意碱基)
5'-GACXALIGGTC-3' 3'-CTGXTAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	PshBI	5'-ATTAAT-3' 3'-TAATTA-5'	5'-XXXX-// -XXATTA-3' 3'-AXXX-// -XTTAA-5' (X为任意碱基)	不含有 5' -TAAT	5'-TAAT	5'-TAATXnTTA (Xn为n个任意碱基)
5'-GACALIIIGGTC-3' 3'-CTGTAACCCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	PshBI	5'-ATTAAT-3' 3'-TAATTA-5'	5'-XXXX-// -XXATTA-3' 3'-AXXX-// -XTTAA-5' (X为任意碱基)	不含有 5' -AAT	5'-AAT	5'-AATXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	AccI Bst1107I	5'-GTATAC-3' 3'-CATATG-5'	5'-XXXX-// -XXGTAA-3' 3'-AXXX-// -XXGATA-5' (X为任意碱基)	不含有 5' -ATAC	5'-ATAC	5'-ATACXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	AccI HincII SalI	5'-GTGAC-3' 3'-CAGCTG-5'	5'-XXXX-// -XXGTGA-3' 3'-AXXX-// -XXGACG-5' (X为任意碱基)	不含有 5' -CGAC	5'-CGAC	5'-CGACXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	AccI Bsp1288I HgaI	5'-GTGAC-3' 3'-CAGCTG-5'	5'-XXXX-// -XXGTGA-3' 3'-AXXX-// -XXGACG-5' (X为任意碱基)	不含有 5' -GCAC	5'-GCAC	5'-GCACXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	HincII HpaI	5'-GTTAAC-3' 3'-CAATTG-5'	5'-XXXX-// -XXGTAA-3' 3'-AXXX-// -XXGATA-5' (X为任意碱基)	不含有 5' -TAAC	5'-TAAC	5'-TAACXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	PsiI	5'-TTATAA-3' 3'-AATATT-5'	5'-XXXX-// -XXTATA-3' 3'-AXXX-// -XXATA-5' (X为任意碱基)	不含有 5' -ATAA	5'-ATAA	5'-ATAAXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	NspV Bsp1104I	5'-TTGAA-3' 3'-AAGCTT-5'	5'-XXXX-// -XXTTGA-3' 3'-AXXX-// -XXGACG-5' (X为任意碱基)	不含有 5' -CGAA	5'-CGAA	5'-CGAAXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	DraI	5'-TTTAAA-3' 3'-AAATTT-5'	5'-XXXX-// -XXTTAA-3' 3'-AXXX-// -XXGATA-5' (X为任意碱基)	不含有 5' -TAAA	5'-TAAA	5'-TAAAXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	ScaI TatI	5'-AGTACT-3' 3'-TCATGA-5'	5'-XXXX-// -XXGATA-3' 3'-AXXX-// -XXTGA-5' (X为任意碱基)	不含有 5' -ACT	5'-ACT	5'-ACTXnTTA (Xn为n个任意碱基)
5'-GACXGLIIGGTC-3' 3'-CTGCAACCCAG-5' (X为任意碱基)	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	SplI	5'-CGTACG-3' 3'-GCATGC-5'	5'-XXXX-// -XXGTA-3' 3'-AXXX-// -XXGTA-5' (X为任意碱基)	不含有 5' -ACG	5'-ACG	5'-ACGXnTTA (Xn为n个任意碱基)

表 6

5'-GACGGTGGTC-3' 3'-CTGCCAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	BspT107 I HgiC I Kpn I	5'-GGTACC-3' 3'-CCATGG-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGCA-5' (X 为任意碱基)	不含有 5' -AGC	5' -ACC	5'-AGCXnTTA (Xn 为 n 个任意碱基)
5'-GACGTTGGTC-3' 3'-CTGCAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	BspI407 I Tat I	5'-TGTACA-3' 3'-ACATGT-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGCA-5' (X 为任意碱基)	不含有 5' -ACA	5' -ACA	5'-ACAXnTTA (Xn 为 n 个任意碱基)
5'-GACACITGGTC-3' 3'-CTGTGAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	Spe I	5'-ACTAGT-3' 3'-TGATCA-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGCA-5' (X 为任意碱基)	不含有 5' -AGT	5' -AGT	5'-AGTXnTTA (Xn 为 n 个任意碱基)
5'-GACCCITGGTC-3' 3'-CTGGCAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	Bln I EcoT14 I	5'-CCTAGG-3' 3'-GGATCC-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGCA-5' (X 为任意碱基)	不含有 5' -AGG	5' -AGG	5'-AGGXnTTA (Xn 为 n 个任意碱基)
5'-GACGCTGGTC-3' 3'-CTGCCAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	Nhe I	5'-GCTAGC-3' 3'-CGATCG-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGCA-5' (X 为任意碱基)	不含有 5' -AGC	5' -AGC	5'-AGCXnTTA (Xn 为 n 个任意碱基)
5'-GACCTCTGGTC-3' 3'-CTGCAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	Xba I	5'-TCTAGA-3' 3'-AGATCT-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGCA-5' (X 为任意碱基)	不含有 5' -AGA	5' -AGA	5'-AGAXnTTA (Xn 为 n 个任意碱基)
5'-GACAAITGGTC-3' 3'-CTGTTAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	Ssp I	5'-AATATT-3' 3'-TTATAA-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGTA-5' (X 为任意碱基)	不含有 5' -ATT	5' -ATT	5'-ATTXnTTA (Xn 为 n 个任意碱基)
5'-GACGALTGGTC-3' 3'-CTGCTAACCAG-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTTACCGACC-5'	EcoRV	5'-GATATC-3' 3'-CTATAG-5'	5'-XXXX-// -XXGTA-3' 3'-AXXXX-// -XXGTA-5' (X 为任意碱基)	不含有 5' -ATC	5' -ATC	5'-ATCXnTTA (Xn 为 n 个任意碱基)

表 7

第一限制酶	第一限制酶的切割识别序列① (含有第二限制酶的切割识别序列的一部分 (下划线部分))	第一限制酶的切割识别序列② (不含第二限制酶的切割识别序列的一部分)	第二限制酶	第二限制酶的切割识别序列	导入到载体中的外来基因 因的5'末端侧和3'末端侧的序列 (大A为在PCR反应中附加的碱基)	用于制作外来基因 的正向引物的5'末端序列应满足的条件	用于制作外来基因 的反向引物的5'末端附加的序列 (不附加终止密码子 的情况)	在用于制作外来基因 的反向引物的5'末端附加的 序列(也包含终止密码子 附加的情况)
XcmI	5'-CCAXCCATGGTATGG-3' 3'-GGTXXGGTACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTGGCGACC-5'	NcoI DsaI EcoII4I	5'-CCATGG-3' 3'-GGTACC-5'	5'-XXXX-//XGCA-3' 3'-AXXXX-//XGCG-5' (X为任意碱基)	不含有5'-GG	5'-GG	5'-GGXnTTA (Xn为n个任意碱基)
	5'-CCAXXCATGGTATGG-3' 3'-GGTXXGATACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTGGCGACC-5'	NdeI	5'-CATATG-3' 3'-GTATAC-5'	5'-XXXX-//XGCA-3' 3'-AXXXX-//XGTA-5' (X为任意碱基)	不含有5'-ATG	5'-ATG	5'-ATGXnTTA (Xn为n个任意碱基)
	5'-CCAXXXCTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	SfeI	5'-CTATAG-3' 3'-GATATC-5'	5'-XXXX-//XGCTATA-3' 3'-AXXXX-//XGCTA-5' (X为任意碱基)	不含有5'-ATAG	5'-ATAG	5'-ATAGXnTTA (Xn为n个任意碱基)
	5'-CCAXCCTATGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	SfeI	5'-CTATAG-3' 3'-GATATC-5'	5'-XXXX-//XGCA-3' 3'-AXXXX-//XGCA-5' (X为任意碱基)	不含有5'-AG	5'-AG	5'-AGXnTTA (Xn为n个任意碱基)
	5'-CCAXCAATGGTATGG-3' 3'-GGTXXGATACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	MunI	5'-CAATGG-3' 3'-GTTAAC-5'	5'-XXXX-//XGCA-3' 3'-AXXXX-//XGCT-5' (X为任意碱基)	不含有5'-TG	5'-TG	5'-TGXnTTA (Xn为n个任意碱基)
	5'-CCAXCGATGGTATGG-3' 3'-GGTXXGATACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	McrI PvuI	5'-CGATCG-3' 3'-GCTAGC-5'	5'-XXXX-//XGCA-3' 3'-AXXXX-//XGCG-5' (X为任意碱基)	不含有5'-CG	5'-CG	5'-CGXnTTA (Xn为n个任意碱基)
	5'-CCAXXXCTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	AvaI SmaI XhoI	5'-CTGAG-3' 3'-GAGCTC-5'	5'-XXXX-//XGCTGCA-3' 3'-AXXXX-//XGAGC-5' (X为任意碱基)	不含有5'-CGAG	5'-CGAG	5'-CGAGXnTTA (Xn为n个任意碱基)
	5'-CCAXXXCTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	PstI SfeI	5'-CTGAG-3' 3'-GAGCTC-5'	5'-XXXX-//XGCTGCA-3' 3'-AXXXX-//XGAGC-5' (X为任意碱基)	不含有5'-GCAG	5'-GCAG	5'-GCAGXnTTA (Xn为n个任意碱基)
	5'-CCAXXXCTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTTATTACCGACC-5'	AflII SmaI	5'-CTTAAG-3' 3'-GAATTC-5'	5'-XXXX-//XGCTTAA-3' 3'-AXXXX-//XGAT-5' (X为任意碱基)	不含有5'-TAAG	5'-TAAG	5'-TAAGXnTTA (Xn为n个任意碱基)

表 8

5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Cla I	5'-ATCGAT-3' 3'-TAGCTA-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -CGAT	5'-CGAT	5'-CGATXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Eco122 I	5'-ATCGAT-3' 3'-TAGCTA-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -GCAT	5'-GCAT	5'-GCATXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Psh I	5'-ATTAAAT-3' 3'-TAATTA-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -TAAT	5'-TAAT	5'-TAATXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Psh I	5'-ATTAAAT-3' 3'-TAATTA-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -AAT	5'-AAT	5'-AATXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Acc I	5'-GTATAC-3' 3'-CATATG-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -ATAC	5'-ATAC	5'-ATACXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Acc I Hinc II Sal I	5'-GTGAC-3' 3'-CAGCTG-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -CGAC	5'-CGAC	5'-CGACXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Bsp1286 I Hga I	5'-GTGAC-3' 3'-CAGCTG-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -GCAC	5'-GCAC	5'-GCACXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Hinc II Hpa I	5'-GTAAAC-3' 3'-CAATTG-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -TAAC	5'-TAAC	5'-TAACXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Psi I	5'-TTATAA-3' 3'-AATATT-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -ATAA	5'-ATAA	5'-ATAAXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Nsp V Bsp1104 I	5'-TTCAA-3' 3'-AAGCTT-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -CGAA	5'-CGAA	5'-CGAAXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Dra I	5'-TTTAA-3' 3'-AAATT-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -TAAA	5'-TAAA	5'-TAAAXnTTA (Xn 为 n 个任意碱基)
5'-CCATTAAGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Sca I Tat I	5'-AGTACT-3' 3'-TCATGA-5'	5'-XXXX-// -XXATGA-3' (X 为任意碱基)	不含有 5' -ACT	5'-ACT	5'-ACTXnTTA (Xn 为 n 个任意碱基)

表 9

5'-CCAXXCGTGGTATGG-3' 3'-GGTXXGACCATACC-5'+B	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Spl I	5'-GGTACG-3' 3'-GCATGC-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -ACG	5'-ACG	5'-ACGnTTA (Xn 为 n 个任意碱基)
5'-CCAXXGGTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Bsp I 107 I Hgi C I Kpn I	5'-GGTACC-3' 3'-CCATGG-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -ACC	5'-ACC	5'-ACCGnTTA (Xn 为 n 个任意碱基)
5'-CCAXXIGTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Bsp I 407 I Tat I	5'-TGTACA-3' 3'-ACATGT-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -ACA	5'-ACA	5'-ACAXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXACTGGTATGG-3' 3'-GGTXXIGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Spe I	5'-ACTAGT-3' 3'-TGAICA-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -AGT	5'-AGT	5'-AGTXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXCGTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Bln I Eco T14 I	5'-CCTAGG-3' 3'-GGATCC-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -AGG	5'-AGG	5'-AGGXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXGCTGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Nhe I	5'-GCTAGC-3' 3'-CGATCG-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -AGC	5'-AGC	5'-AGCXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXICTGGTATGG-3' 3'-GGTXXAGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Xba I	5'-TCTAGA-3' 3'-AGATCT-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -AGA	5'-AGA	5'-AGAXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXAATGGTATGG-3' 3'-GGTXXIIACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Ssp I	5'-AATATT-3' 3'-TTATAA-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -ATT	5'-ATT	5'-ATTXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXGATGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	EcoRV	5'-GATATC-3' 3'-CTATAG-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -ATC	5'-ATC	5'-ATCXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXAATGGTATGG-3' 3'-GGTXXIIACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Apo I	5'-AAATTT-3' 3'-TTTAAA-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -TTT	5'-TTT	5'-TTXnTTA (Xn 为 n 个任意碱基)
5'-CCAXXGATGGTATGG-3' 3'-GGTXXGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	Apo I EcoRI	5'-GAATTC-3' 3'-CTAAG-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -TC	5'-TC	5'-TCXnTTA (Xn 为 n 个任意碱基)
5'-CCAXACATGGTATGG-3' 3'-GGTXXIGACCATACC-5'	5'-CCAATAAATGGCTGG-3' 3'-GGTATTATCCGACC-5'	AflIII BspLU1 I Nsp I	5'-ACATGT-3' 3'-TGACA-5'	5'-XXXX-// -XNCTA-3' 3'-AXXXX-// -XNCA-5' (X 为任意碱基)	不含有 5' -GT	5'-GT	5'-GTXnTTA (Xn 为 n 个任意碱基)

表 10

5'-CCAXGGATGGTATGG-3' 3'-GGTXXGATACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Nsp I Sph I	5'-GCATGC-3' 3'-GCTACG-5'	5'-XXXX-// -XXXX-3' 3'-AXXXX-// -XXCC-5' (X 为任意碱基)	不含有 5' -GC	5'-GC	5'-GCXnTTA (Xn 为 n 个任意碱基)
5'-CCAXICATGGTATGG-3' 3'-GGTXGATACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Esp HI	5'-TCATGA-3' 3'-AGTACT-5'	5'-XXXX-// -XTTCA-3' 3'-AXXXX-// -XXAG-5' (X 为任意碱基)	不含有 5' -GA	5'-GA	5'-GAXnTTA (Xn 为 n 个任意碱基)
5'-CCAXAGATGGTATGG-3' 3'-GGTXXCTACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Bcl II Mfi I	5'-AGATCT-3' 3'-TCTAGA-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXTC-5' (X 为任意碱基)	不含有 5' -CT	5'-CT	5'-CTXnTTA (Xn 为 n 个任意碱基)
5'-CCAXGGATGGTATGG-3' 3'-GGTXXCTACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Eam HI Mfi I	5'-GGATCC-3' 3'-CCTAGG-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXCC-5' (X 为任意碱基)	不含有 5' -CC	5'-CC	5'-CCXnTTA (Xn 为 n 个任意碱基)
5'-CCAXIGATGGTATGG-3' 3'-GGTXACTACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Fba I	5'-TGATCA-3' 3'-ACTAGT-5'	5'-XXXX-// -XTTCA-3' 3'-AXXXX-// -XXCC-5' (X 为任意碱基)	不含有 5' -CA	5'-CA	5'-CAXnTTA (Xn 为 n 个任意碱基)
5'-CCAXGIATGGTATGG-3' 3'-GGTXCATACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Acc I Bst1107 I	5'-GTATAC-3' 3'-CATATG-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXCA-5' (X 为任意碱基)	不含有 5' -AC	5'-AC	5'-ACXnTTA (Xn 为 n 个任意碱基)
5'-CCAXIIATGGTATGG-3' 3'-GGTXAATACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Psi I	5'-TTATAA-3' 3'-AATATT-5'	5'-XXXX-// -XTTCA-3' 3'-AXXXX-// -XXAA-5' (X 为任意碱基)	不含有 5' -AA	5'-AA	5'-AAXnTTA (Xn 为 n 个任意碱基)
5'-CCAAAGCTGGTATGG-3' 3'-GGTIIAGCAACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	PspI 406 I	5'-AACGTT-3' 3'-TTGCAA-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXCT-5' (X 为任意碱基)	不含有 5' -T	5'-T	5'-TXnTTA (Xn 为 n 个任意碱基)
5'-CCACAGCTGGTATGG-3' 3'-GGTGTGCAACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Bsa AI PmeC I	5'-CACGTG-3' 3'-GTGCAC-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXCG-5' (X 为任意碱基)	不含有 5' -G	5'-G	5'-GXnTTA (Xn 为 n 个任意碱基)
5'-CCAGAGCTGGTATGG-3' 3'-GGTCTGCACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Aat II Hinf I	5'-GACGTC-3' 3'-CTGCAG-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXCC-5' (X 为任意碱基)	不含有 5' -C	5'-C	5'-CXnTTA (Xn 为 n 个任意碱基)
5'-CCATACGTGGTATGG-3' 3'-GGTATGCAACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Bsa AI SnaBI	5'-TACGTA-3' 3'-ATGCAAT-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXCA-5' (X 为任意碱基)	不含有 5' -A	5'-A	5'-AXnTTA (Xn 为 n 个任意碱基)
5'-CCAAGCTGGTATGG-3' 3'-GGTIIAGCAACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTTGCCGACC-5'	Hind III	5'-ARGCTT-3' 3'-TTCGAA-5'	5'-XXXX-// -XXXXA-3' 3'-AXXXX-// -XXCT-5' (X 为任意碱基)	不含有 5' -T	5'-T	5'-TXnTTA (Xn 为 n 个任意碱基)

表 11

5'-CCACAGCTGGTATGG-3' 3'-GGTGTGACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTGGCGACC-5'	NspII PvuII	5'-CAGCTG-3' 3'-GTGAC-5'	5'-xxxx-// -xxxxA-3' 3'-Axxxx-// -xxxG-5' (X 为任意碱基)	不含有 5' -G	5'-G	5'-GXnTTA (Xn 为 n 个任意碱基)
5'-CCAGAGCTGGTATGG-3' 3'-GGTGTGACCATACC-5'	5'-CCAATAAACGGCTGG-3' 3'-GGTTATTGGCGACC-5'	BanII BspI 286 I HgaI SacI	5'-GAGCTC-3' 3'-CTCGAG-5'	5'-xxxx-// -xxxxA-3' 3'-Axxxx-// -xxxG-5' (X 为任意碱基)	不含有 5' -C	5'-C	5'-CXnTTA (Xn 为 n 个任意碱基)

当第一限制性酶为 Ahd I 或其同裂酶、Xcm I 或其同裂酶 (isoschizomer)、或它们的组合时,当第二限制性酶为 Nco I、Nde I 或其同裂酶时,是有效的。

通过利用本发明的载体,可以制作如下所述的文库 (library) (例如
5 参照图 3),即设计例如混合引物,使用该混合引物并将目的蛋白质的 cDNA 用作模板使其扩增,将编码任意蛋白质的部分序列的 DNA 片断的混合物连接于其它蛋白质的下游,能够使原来的蛋白质中位置和/或长度不同的部分序列的融合蛋白质表达。

另外,通过利用本发明的载体,例如可以使用 96 孔微滴定板
10 (microtitre plate),网罗地制作多个蛋白质和/或蛋白质结构域的表达系统,而且网罗地检测利用该表达系统产生的蛋白质和/或蛋白质结构域的溶解度,或者高速地只选择出溶解度高的蛋白质和/或蛋白质结构域(例如,参照图 3)。

另外,使用本发明的载体和 96 孔微滴定板的上述方法,通过使用自
15 动菌落分离装置、或能够操作 96 孔微滴定板的自动试剂分注清洗装置、实验室自动化 (laboratory automation) 装置,能够网罗地制作多种蛋白质和/或蛋白质结构域的表达系统,而且能够全自动并高速地选别利用该表达系统产生的溶解度高的蛋白质和/或蛋白质结构域。

进而,通过利用本发明的载体,将单一蛋白质和/或蛋白质结构域作
20 为对象,可以在从 PCR 反应开始最短 48 小时内进行如下所述的判别:判断含有该单一蛋白质和/或蛋白质结构域的融合蛋白质是否可以在大肠杆菌中表达,以及判断含有该单一蛋白质和/或蛋白质结构域的融合蛋白质是否为可溶性(例如,参照图 7)。

25 实施例

下面,通过实施例对本发明进行具体地说明。其中,这些实施例是用于说明本发明的实施例,不限定本发明的范围。

限制性酶 Nco I、Nde I 以及 Afl II 是从 New England Biolabs 公司购入。Kpn I 和 BamH I 是从东洋纺公司购入。T4 DNA 连接酶 (ligase)、
30 Wizard 质粒小量制备试剂盒 (plasmid miniprep kit) 以及 Wizard SV DNA

凝胶纯化试剂盒 (gel purification kit) 是从 Promega 公司购入。pGEM-T、pGEX-4T3 和 pGEX-3X、pET32a 和 pET21b 分别是 Promega 公司、Amersham Pharmacia Biotech 公司以及 Novagen 公司购入。另外, pET32a PACAP 是在 pET21b 的 Nde I 位点插入含有 pET32a 来源的硫氧还蛋白 (thioredoxin) 基因的 Nde I 片断和在该 pET21b 的 Nhe I 位点插入 PACAP 的人工基因而作成的 pET32a 的衍生物。作为用作 PCR 克隆的模板的 cDNA, 小鼠脾脏、小鼠胎儿以及人 HeLa 细胞的 QUICK-Clone cDNA 是从 Clontech 公司购入。

[实施例 1] dGST-T 载体的制作 (图 1)

使用对序列号 21 的序列所表示的 DNA 进行 5'磷酸化而成的产物, 利用定位突变法, 向 pGEX-4T3 载体导入突变。这样, pGEX-4T3 在原来的序列中具有 Ahd I 限制性酶位点不会被切割。接着, 将 pET32aPACAP 载体作为模板, 以用序列号 22 和 24 的序列表示的引物的对进行 PCR 反应。得到的 PCR 产物 (大约 500bp, 将双链 DNA 的一条链的序列表示在序列号 25 中) 分别在两个 BamH I 部位结合 Ahd I 部位 (图 1)。将该 PCR 产物亚克隆到 TA 克隆载体 pGEX-T 中, 在确认其序列之后, 制备纯化质粒。用限制性酶 BamH I 切割纯化后的质粒, 纯化约 500 碱基的片断 (Ahd I 连接序列, 序列号 26 和 27)。用 BamH I 使上述已导入突变的 pGEX-4T3 开环, 实施脱磷酸化处理, 使得到的产物与上述 Ahd I 连接序列连接, 转化大肠杆菌。确认序列, 然后选择已将 Ahd I 连接序列以正向导入到 GST 基因下游的载体。用 25 单位的 Ahd I 限制性酶在 37℃ 下处理已将 Ahd I 连接序列正向导入到 GST 基因下游的载体 (约 2μg) 3 小时并切割, 使用 1% 琼脂糖凝胶 (agarose gel) 进行纯化。使用 Wizard SV 凝胶/PCR 清洗 (clean-up) 系统纯化约 5kb 的开环式载体 (pGEX-4T3 PRESAT, 图 1), 用于下述实施例中。

[实施例 2] 利用新型表达系统的并行操作的制备 (图 2)

首先, 利用 Nco I 一选择法对应的引物, 同时并行将人、小鼠等的 cDNA 作为模板, 用 PCR 扩增五种基因 (含有人、小鼠的 SNX15a 来源以及酵母、小鼠、人来源的 Vps4b 的五个 MIT 结构域的区域, 序列号 28、30、32、34、36)。使用的引物 (序列号 38~47) 是, 与在该基因的 cDNA

中预计作为蛋白质表达的区域互补的序列、终止密码子的互补序列、以及被适当设计成在 5' 末端含有用于进行 Nco I 选择的 GG 残基的序列。即，正向引物以最初的氨基酸的密码子开始并与前方的 GST 为读码框一致，5' 末端的碱基不是从 5' -GG 开始。反向引物含有该结构域的推测 C 末端边界后的终止密码子，5' 末端的碱基必须从 5' -GG 开始。这样，以这些引物扩增出的 PCR 产物只有在被逆向连接到 pGEX-4T3 PRESAT 时才形成第二个 Nco I 部位。该新产生的 Nco I 部位被用于接下来的选择工序中。

该基因均为 200~230bp 的 cDNA，以新的蛋白质结构域编码约 70 个氨基酸残基。

在通过电泳确认 PCR 产物正在扩增之后，在未纯化的状态下使该产物与 pGEX-4T3 PRESAT 载体发生连接反应，转化大肠杆菌 DH5a。转化后的大肠杆菌 DH5a 在不进行单菌落化的情况下进行整夜液体培养，次日从中纯化出 1~3 μ g 的质粒。将用 Nco I 限制性酶切割该质粒的产物转化表达用大肠杆菌宿主 BL21 (DE3)，进行单菌落化。从该已菌落化的板分别取三个菌落 ($3 \times 5 = 15$)，用 LB 培养基进行液体培养，在接近 OD=0.5 时加入 IPTG 并使其终浓度为 1mM，进一步培养 4 个小时。收集表达结束后的大肠杆菌，超声波破碎后，通过 SDS-PAGE 使全菌体泳动，在凝胶上确认目的融合蛋白质的带 (约 33~35kD) (图 4) (表达性试验)。

在得到的菌落中，对预测分子量的目的融合蛋白质进行表达的菌落的比例约 90% 以上 (14/15)，效率非常高。

进而，通过离心分离，将已确认了目的分子量的融合蛋白质的表达的菌落来源的菌体破碎液分离为可溶性组分、不溶性组分，与全菌体一起，用 SDS-PAGE 分别泳动，检验目的融合蛋白质是否为可溶性 (图 5) (可溶性试验)。

另外，向已确认目的分子量的融合蛋白质的表达的菌落来源的蛋白质的可溶性组分的一部分中，加入由 1-氯-2、4-二硝基苯和谷胱甘肽的混合液构成的 GST 活性检测液，开始基于 GST 的反应，用分光光度计测量 350nm 的紫外部分吸光度的增大来计量其反应速度 (图 6) (CDNB 检测)。曲线的直线斜率被确认为与目的融合蛋白质的可溶性组分中的量大

致成比例，由此可知利用分光光度计的半定量可溶性试验是可能的。

与此相独立，用菌落 PCR，对同一板的菌落各八个，探测各菌落中含有的质粒中含有目的基因的比例(亚克隆的效率)。为 37/40 的概率(92.5 %)，确认了已结合从靶基因可以预测的正确长度的基因。

- 5 从表达正确长度的蛋白质的克隆选择两个，判定确认 DNA 碱基序列。已确认序列全部正确，另外与 GST 无读码框错位。

以上操作全部工序需要的天数为约六天，其缩短有意义。

[实施例 3] Nco I 选择法与 Nde I 选择法的效率比较

- 在制作含有与泛素(ubiquitin)相互作用的基元(motif)(UIM:
10 ubiquitin interacting motif)的蛋白质片断的 GST 融合蛋白质表达系统时，通过 Nco I 选择与 Nde I 选择，进行比较实验以确认实施上的表达系统构建的效率是否不同。

- 利用 Nco I 一选择法对应的引物的组(序列号 48 和 45)以及 Nde I 选择法对应的引物的组(序列号 48 和 50)，用 PCR，从人的 cDNA 扩增
15 含有 UIM 序列的人 Hrs 基因的约 200bp 的片断(将编码链的序列显示于序列号 51)。使用的引物是，与该基因的 cDNA 互补的序列、终止密码子的互补序列、以及被适当设计成在 5' 末端含有 GG 残基以进行 Nco I 选择的序列。用于进行 Nde I 选择的引物是，终止密码子的互补序列、以及被适当设计成在 5' 末端含有 ATG 残基的序列。该片断为 200bp 的 cDNA，
20 以新的蛋白质结构域编码约 70 个氨基酸残基。

- 在利用电泳确认 PCR 产物正在扩增之后，在未纯化的情况下使该产物与 dGST-T 载体发生连接反应，转化大肠杆菌 DH5a。转化后的大肠杆菌 DH5a 在不进行单菌落化的情况下进行整夜液体培养，次日从中纯化出 1~3 μ g 的质粒。将用 Nco I 限制性酶或 Nde I 限制性酶切割该质粒的产物转化表达用大肠杆菌宿主 BL21 (DE3)，单菌落化。反应条件为质粒
25 (plasmid) 200ng、使用 10 单位(unit) Nco I、20 单位 Nde I，37℃下均反应一个小时。从该已菌落化的板分别取三个菌落(3×5=15)，用 LB 培养基进行液体培养，在 OD=0.5 左右时加入 IPTG 并使其终浓度为 1mM，进一步培养 4 个小时(表达试验)。收集表达结束后的大肠杆菌，
30 超声波破碎后，通过 SDS-PAGE 使全菌体、可溶性组分、不溶性组分各

自泳动，在凝胶上确认目的融合蛋白质的带（约 31kD）。在得到的菌落中，对预测分子量的目的融合蛋白质进行表达的菌落的比例在 Nco I 选择时为 14/15、在 Nde I 选择时为 3/3，均为约 90%以上（14/15），效率非常高。

与此相独立，用菌落 PCR，对同一板的菌落的数个，探测各菌落中所含的质粒中含有目的基因的比例（亚克隆的效率）。已确认在 Nco I 选择时为 9/10 的概率（90.0%）、在 Nde I 选择时为 7/8（88%），且以大致相同的效率结合了从靶基因预测的正确长度的基因。

从以上实验可知，Nco I / Nde I 均可以用于 ORF 方向的选择。

[实施例 4] 在短时间内对利用 dGST-T 载体制作而成的融合蛋白质在大肠杆菌中的表达量和溶解度进行试验的方法（图 7）

使用已插入在实施例 2 中制作的人 Vps4b 的基因的且通过 Nco I 处理的 ORF 选择已终止的开环状载体混合物约 0.2 μ g，确定用于在短时间内进行其中被认为是可以完成的 GST-Vps4b 的表达性试验和溶解度试验的条件。

将约 0.2 μ g 的已经在实施例 2 中记载的且已经用 Nco I 处理成为开环状的、含有人 Vps4b 的基因的 pGEX-4T3-PRESAT 转化表达用大肠杆菌宿主 BL21 (DE3)，使其在 LB 培养基上、在 37℃下用 14 小时形成菌落。在已形成该菌落的板上，喷雾 0.1M 的 IPTG 溶液 200 μ l，进一步在 37℃下使菌落成长 3 个小时，在板上表达蛋白质。向板上的大肠杆菌菌落加入 180 μ l 的含有 0.15M NaCl 的 20mM Tris 盐酸缓冲液（pH 7.0）（以下称为 TBS 缓冲液），悬浮，超声波破碎该菌体，与实施例 2 一样，通过 SDS-PAGE 电泳分析全菌体组分、可溶性组分、不溶性组分（图 8）。从大肠杆菌的悬浮进行利用 SDS-PAGE 的分析所需要的时间约为 3 个小时。

在该方法中，也与实施例 2 相同，在 SDS-PAGE 上确认作为目的的分子量约为 32KDa 的融合蛋白质的带，可知表达量和可溶性在今后的实验中使用是足够的。

在该实验中，没有使含有目的融合蛋白质的大肠杆菌单克隆化，而是以混合物形式进行表达性试验和溶解度试验，但如实施例 2 中所示，在进行利用 Nco I 的 ORF 选择的情况下，不含有目的融合蛋白质的大肠杆菌

为在板培养基上形成菌落的大肠杆菌的 10% 以下，所以可知利用通过电泳的表达性试验、可溶性试验进行判断是可能的。

以上的操作全部工序需要的时间为大约 21 小时，缩短有意义。从作为其前一阶段的 PCR 开始到利用 Nco I 的 ORF 处理所需要的时间最短为
5 27 小时是可能的，以最短 48 小时进行单一蛋白质和/或蛋白质结构域的表达系统构建和可溶性试验的整个工序是可能的。

本发明人等制作的载体的名称、成为载体制作的基础的市售质粒的名称、标签 (tag) 的名称、第一限制性酶、第二限制性酶、连接序列的序列以及参考文献被归纳于下述表中。

表 12

载体名	成为载体制作 的基础的市售 质粒的名称	标签的名称 *	第一限 制性酶	第二限 制性酶	连接序列 的序列	参考 文献
pGEX-4T3-PRESAT	pGEX-4T3	GST-凝血酶切割位点	AhdI	NcoI NdeI	I	1, 7
pGEX-3X-PRESAT	pGEX-3X	GST-因子 Xa 切割位点	AhdI	NcoI NdeI	I	1, 7
pET-Trx-PRESAT	pET-32a	TRX-His6-因子 Xa 切割位点	AhdI	KpnI	II	2, 3, 8
pET-Hisx6 -PRESAT	pET-32a	His6-因子 Xa 切割位点	AhdI	KpnI	II	8
pET-FLAG-PRESAT	pET-32a	FLAG (注册商标) -肠激酶切割位点	AhdI	AflII	III	4
pET-LBT-PRESAT	pET-32a	LBT	AhdI	KpnI	II	5
pET-EGFP-PRESAT	pET-32a	EGFP-His6	AhdI	KpnI	II	6

*包括标签以及切开标签和目的蛋白质时的蛋白酶切割序列的名称简写

GST: 谷胱甘肽-S-转移酶 (Schistosoma japonicum)

TRX: 硫氧还蛋白 (大肠杆菌 (E. coli))

5 His6: His-His-His-His-His-His

LBT: 镧系金属结合标签 (lanthanide binding tag)

FLAG (注册商标): Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys

EGFP: 增强绿色荧光蛋白 (enhanced green fluorescence protein) (多管水母属 (オワンクラゲ))

10

连接序列的序列 (下划线部分的序列为用 Ahd I 的切割位点)

I (487bp) 说明: Trx 基因与-PACAP 人工基因的一部分

GACCATTGGTCCACCTGACTGACGACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGC

GATCCTCGTCGATTTCTGGGCAGAGTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTG

15 GATGAAATCGCTGACGAATATCAGGGCAAAGTACCGTTGCAAACTGAACATCGATCAAAA

CCCTGGCACTGCGCCGAAATATGGCATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAAC

GGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCG

ACGCTAACCTGGCCGGTTCTGGTTCTGGCCATATGGCTAGCCATCACCACCACCACCACAG

CAGCGGCATTGACGGCCGGCATAGCGATGGCATCTTTACCGATAGCTATAGCCGCTATCGC

20 AAACAGATGGCGGTGAAAAAGTATCTGGCGGCGGTGCTGGGCTAATAAGACCAAATGTC (序列号 59)

II (110bp) 说明: PACAP 人工基因

GACGGTCGGTCTAGCGATGGCATCTTTACCGATAGCTATAGCCGCTATCGCAAACAGATGGC

25 GGTGAAAAAGTATCTGGCGGCGGTGCTGGGCTAATAAGACCAATGGTC (序列号 60)

III (259bp) 说明: 人泛素基因

GACGATAAGTCTCATATGCAGATTTTCGTGAAAACCCCTGACCGGCAAAACCATCACCTGGA
AGTGGAACCGTCCGACACCATCGAAAACGTTAAAGCGAAAATCCAGGACAAAGAAGGCATC
CCGCCGGATCAGCAGCGTCTGATCTTCGCGGGCAAACAGCTGGAAGACGGCCGTACCCTG
TCCGATTACAACATCCAGAAAGAATCTACCCTGCACCTGGTGCTGCGTCTGCGTGGCGGCT
5 AATAGACTTAAGGTC (序列号 61)

参考文献

- 1 Smith DB, Johnson KS. (1998) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* 67, 31-40
- 2 LaVallie, E. R, DiBlasio, E. A, Kovacic, S, Grant, K. L, Schendel, P. F. and McCoy, J. M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio/Technology* 11, 187-193
- 3 Novy, R, Berg, J, Yaeger, K, and Mierendorf, R. (1995) *inNovations* 3, 7-9
- 4 Hopp TP, Prickett KS, Price V, Libby RT, March CJ, Cerretti P, Urdal DL, and Conlon PJ. (1988) A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Biotechnology* 6, 1205-1210
- 5 Wohnert J, Franz KJ, Nitz M, Imperiali B, Schwalbe H. (2003) Protein Alignment by a Coexpressed lanthanide-Binding Tag for the Measurement of Residual Dipole Couplings. *J. Am. Chem. Soc.* 125,

13338-13339

6 Waldo GS, Standish BM, Berendzen J, Terwilliger TC. (1999)
Rapid protein-folding assay using green fluorescent protein. *Nat*
Biotechnol. 17, 691-695.

5 7 Goda N, Tenno T, Takasu H, Hiroaki H, and Shirakawa M. (2004)
The PRESAT-vector: Asymmetric T-vector for high-throughput
screening of soluble protein domains for structural proteomics.
Protein Science 13, 652-658

8 Tenno T, Goda N, Tateishi Y, Tochio H, Mishima M, Hayashi H,
10 Shirakawa M, and Hiroaki H. (2004) High-throughput construction
method for expression vector of peptides for NMR study suited for
isotopic labeling. *Protein Engineering, Design and Selection* 17,
305-314

15 在本说明书中引用的全部刊物、专利以及专利申请被直接作为参考，
列入本说明书中。

工业上的可利用性

通过本发明，可以大幅度缩短在蛋白质和/或蛋白质结构域的表达系
20 统构建和可溶性试验的全部工序中所需要的时间。

序列表 (free text)

<序列号 1>

序列号 1 表示序列 (I) 的碱基序列。

25 <序列号 2>

- 序列号 2 表示序列 (II) 的碱基序列。
<序列号 3>
序列号 3 表示序列 (III) 的碱基序列。
<序列号 4>
5 序列号 4 表示序列 (IV) 的碱基序列。
<序列号 5>
序列号 5 表示序列 (V) 的碱基序列。
<序列号 6>
序列号 6 表示序列 (VI) 的碱基序列。
10 <序列号 7>
序列号 7 表示序列 (VII) 的碱基序列。
<序列号 8>
序列号 8 表示序列 (VIII) 的碱基序列。
<序列号 9>
15 序列号 9 表示序列 (IX) 的一个例子 (N^1 为一个碱基的情况) 的碱基序列。
<序列号 10>
序列号 10 表示序列 (X) 的一个例子 (N^2 为一个碱基的情况) 的碱基序列。
20 <序列号 11>
序列号 11 表示序列 (XI) 的一个例子 (N^1 为一个碱基的情况) 的碱基序列。
<序列号 12>
序列号 12 表示序列 (XII) 的一个例子 (N^2 为一个碱基的情况) 的碱基序列。
25 列。
<序列号 13>
序列号 13 表示序列 (XIII) 的一个例子 (N^1 为一个碱基的情况) 的碱基序列。
<序列号 14>
30 序列号 14 表示序列 (XIV) 的一个例子 (N^2 为一个碱基的情况) 的碱基

序列。

<序列号 15>

序列号 15 表示序列 (X V) 的一个例子 (N^3 为一个碱基的情况) 的碱基序列。

5 <序列号 16>

序列号 16 表示序列 (X VI) 的一个例子 (N^4 为一个碱基的情况) 的碱基序列。

<序列号 17>

10 序列号 17 表示序列 (X VII) 的一个例子 (N^3 为一个碱基的情况) 的碱基序列。

<序列号 18>

序列号 18 表示序列 (X VIII) 的一个例子 (N^4 为一个碱基的情况) 的碱基序列。

<序列号 19>

15 序列号 19 表示序列 (X IX) 的一个例子 (N^3 为一个碱基的情况) 的碱基序列。

<序列号 20>

序列号 20 表示序列 (X X) 的一个例子 (N^4 为一个碱基的情况) 的碱基序列。

20 <序列号 21>

序列号 21 表示用于 pGEX-4T3 的部位特异突变法中的 DNA 的序列。

<序列号 22>

序列号 22 表示在将 pET32aPACAP 作为模板的 PCR 反应中使用的引物(正向)的碱基序列。

25 <序列号 23>

序列号 23 表示在将 pET32aPACAP 作为模板的 PCR 反应中使用的引物(反向)的碱基序列。

<序列号 24>

30 序列号 24 表示在将 pET32aPACAP 作为模板的 PCR 反应中使用的引物 (Nco I 选择法对应的反向) 的碱基序列。

<序列号 25>

序列号 25 表示实施例 1 的约 500bp 的 PCR 产物的一条 DNA 链的碱基序列。另一条 DNA 链是其互补链。

<序列号 26>

- 5 序列号 26 表示实施例 1 的作为 Ahd I 连接序列的双链 DNA 的一条链的碱基序列。

<序列号 27>

序列号 27 表示实施例 1 的作为 Ahd I 连接序列的双链 DNA 的另一条链的碱基序列。

10 <序列号 28>

序列号 28 表示成为 PCR 的模板的、含有 MIT 结构域的基因(人 SNX15a)的约 220bp 的片断(编码链)的碱基序列。不含有通过引物附加的序列。

<序列号 29>

- 15 序列号 29 表示由成为 PCR 的模板的、含有 MIT 结构域的基因(人 SNX15a)的约 220bp 的片断(编码链)所编码的氨基酸序列。

<序列号 30>

序列号 30 表示成为 PCR 的模板的、含有 MIT 结构域的基因(小鼠 SNX15a)的约 220bp 的片断(编码链)的碱基序列。不含有通过引物附加的序列。

<序列号 31>

- 20 序列号 31 表示由成为 PCR 的模板的、含有 MIT 结构域的基因(小鼠 SNX15a)的约 220bp 的片断(编码链)所编码的氨基酸序列。

<序列号 32>

序列号 32 表示成为 PCR 的模板的、含有 MIT 结构域的基因(酵母 Vps4)的约 220bp 的片断(编码链)的碱基序列。不含有通过引物附加的序列。

25 <序列号 33>

序列号 33 表示由成为 PCR 的模板的、含有 MIT 结构域的基因(酵母 Vps4)的约 220bp 的片断(编码链)所编码的氨基酸序列。

<序列号 34>

- 30 序列号 34 表示成为 PCR 的模板的、含有 MIT 结构域的基因(人 Vps4b)的约 220bp 的片断(编码链)的碱基序列。不含有通过引物附加的序列。

<序列号 35>

序列号 35 表示由成为 PCR 的模板的、含有 MIT 结构域的基因(人 Vps4b)的约 220bp 的片断(编码链)所编码的氨基酸序列。

<序列号 36>

- 5 序列号 36 表示成为 PCR 的模板的、含有 MIT 结构域的基因(小鼠 Vps4b)的约 220bp 的片断(编码链)的碱基序列。不含有通过引物附加的序列。

<序列号 37>

序列号 37 表示由成为 PCR 的模板的、含有 MIT 结构域的基因(小鼠 Vps4b)的约 220bp 的片断(编码链)所编码的氨基酸序列。

- 10 <序列号 38>

序列号 38 表示引物(HSS15F)的碱基序列。

<序列号 39>

序列号 39 表示引物(HSS15R)的碱基序列。

<序列号 40>

- 15 序列号 40 表示引物(MMS15F)的碱基序列。

<序列号 41>

序列号 41 表示引物(MMS15R)的碱基序列。

<序列号 42>

序列号 42 表示引物(ScVPS4F)的碱基序列。

- 20 <序列号 43>

序列号 43 表示引物(ScVPS4R)的碱基序列。

<序列号 44>

序列号 44 表示引物(HsVPS4F)的碱基序列。

<序列号 45>

- 25 序列号 45 表示引物(HsVPS4R)的碱基序列。

<序列号 46>

序列号 46 表示引物(MmVPS4F)的碱基序列。

<序列号 47>

序列号 47 表示引物(MmVPS4R)的碱基序列。

- 30 简称 Sc: 酵母(yeast) Hs: 人 Mm: 小鼠: VPS4: VPS4, S15: SNX15 F:

正向 (FORWARD) R:反向 (REVERSE)

<序列号 48>

序列号 48 表示引物 (HrsF) 的碱基序列。

<序列号 49>

- 5 序列号 49 表示 Nco I 选择对应的引物 (HrsR-Nco) 的碱基序列。

<序列号 50>

序列号 50 表示 Nde I 选择对应的引物 (HrsR-Nde) 的碱基序列。

<序列号 51>

- 10 序列号 51 表示成为 PCR 的模板的、含有 UIM 序列的人 Hrs 基因的约 200bp 的片断 (编码链) 的碱基序列。不含有通过引物附加的序列。

<序列号 52>

序列号 52 表示 由成为 PCR 的模板的、含有 UIM 序列的人 Hrs 基因的约 200bp 的片断 (编码链) 所编码的氨基酸序列。

<序列号 53>

- 15 序列号 53 表示当第一限制性酶为 Ahd I、第二限制性酶为 Nco I 时的 ORF 的 C 末端侧的引物的碱基序列 (目的蛋白质结构域不含有终止密码子时)。

<序列号 54>

序列号 54 表示当第一限制性酶为 Ahd I、第二限制性酶为 Nde I 时的 ORF 的 C 末端侧的引物的碱基序列 (目的蛋白质结构域不含有终止密码子时)。

- 20 <序列号 55>

序列号 55 表示序列 (IX) ~ (XIV) 中的 N¹ 的碱基序列的一个例子。

<序列号 56>

序列号 56 表示序列 (IX) ~ (XIV) 中的 N² 的碱基序列的一个例子。

<序列号 57>

- 25 序列号 57 表示序列 (XV) ~ (XX) 中的 N³ 的碱基序列的一个例子。

<序列号 58>

序列号 58 表示序列 (XV) ~ (XX) 中的 N⁴ 的碱基序列的一个例子。

<序列号 59>

序列号 59 表示连接序列 I 的碱基序列。

- 30 <序列号 60>

序列号 60 表示连接序列 II 的碱基序列。

<序列号 61>

序列号 61 表示连接序列 III 的碱基序列。

<110> 横滨市

<120> 新型载体及其利用

<130> FP-038PCT

<140>

<141>

<150>JP P2003-308773

<151>2003-09-01

<160> 61

<170> PatentIn Ver. 2.1

<210> 1

<211> 11

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述: 合成 DNA

<400> 1

gacnntnngt c

11

<210> 2

<211> 11

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述: 合成 DNA

<400> 2

gacnnanngt c

11

<210> 3

<211> 11

<212> DNA

<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 3
gacnnanngt c 11

<210> 4
<211> 11
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 4
gacnntnngt c 11

<210> 5
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 5
ccannnnntnn nntgg 15

<210> 6
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 6
ccannnnnann nntgg 15

<210> 7
<211> 15
<212> DNA
<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 7

ccannnnann nntgg

15

<210> 8

<211> 15

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 8

ccannntnn nntgg

15

<210> 9

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 9

gacnntnngt cngacnnann gtc

23

<210> 10

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 10

gacnntnngt cngacnnann gtc

23

<210> 11

<211> 31

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 11

ccannntnn nntggnccan nnnannntg g

31

<210> 12

<211> 31

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 12

ccannntnn nntggnccan nnnannntg g

31

<210> 13

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 13

gacnntnngt cnccannna nnnntgg

27

<210> 14

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 14

ccannntnn nntgngacn nanngtc

27

<210> 15

<211> 12

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 15

nngtcngacn nt

12

<210> 16

<211> 12

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 16

nngtcngacn nt

12

<210> 17

<211> 16

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 17

nnnntgggcc annnnt

16

<210> 18

<211> 16

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 18

nnnntgggcc annnnt

16

<210> 19

<211> 14

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 19

nnnttgnga cntt

14

<210> 20

<211> 14

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 20

nngtcncaan nnnt

14

<210> 21

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 21

gttatctaca ccacggggag ccaggcaact atgg

34

<210> 22

<211> 41

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 22

ggatccgacc attggtccac ctgactgacg acagttttga c

41

<210> 23

<211> 37

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 23

ggatccgacc attggtctta ttagcccagc accgccg

37

<210> 24

<211> 37

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 24

ggatccgaca tttggtctta ttagcccagc accgccg

37

<210> 25

<211> 499

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 25

```

ggatccgacc attggtccac ctgactgacg acagttttga cacggatgta ctcaaagcgg 60
acggggcgat cctcgtcgat ttctgggcag agtgggtcgg tccgtgcaaa atgatcgccc 120
cgattctgga tgaaatcgct gacgaatatc agggcaaaact gaccgttgca aaactgaaca 180
tcgatcaaaa ccctggcact ggcgcgaaat atggcatccg tggatcccg actctgctgc 240
tggtcaaaaa cgggtgaagtg gcggcaacca aagtgggtgc actgtctaaa ggtcagttga 300
aagagttcct cgacgctaac ctggccggtt ctggttctgg ccatatggct agccatcacc 360
accaccacca cagcagcggc attgacggcc ggcatagcga tggcatcttt accgatagct 420
atagccgcta tcgcaaacag atggcgggtga aaaagtatct ggccggcgggtg ctgggctaata 480
aagaccaaata gtcggatcc 499

```

<210> 26

<211> 493

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 26

```

gatccgacca ttggtccacc tgactgacga cagttttgac acggatgtac tcaaagcgga 60
cggggcgatc ctctcgatt tctgggcaga gtggtgoggt ccgtgcaaaa tgatcgcccc 120
gattctggat gaaatcgctg acgaatatca gggcaaaactg accgttgcaa aactgaacat 180
cgatcaaaac cctggcactg cgccgaaata tggcatccgt ggtatccga ctctgctgct 240
gttcaaaaac ggtgaagtgg cggcaaccaa agtgggtgca ctgtctaaag gtcagtigaa 300
agagttcctc gacgctaacc tggccgggtc tggttctggc catatggcta gccatcacca 360
ccaccaccac agcagcggca ttgacggcgg gcatacgcat ggcattttta ccgatagcta 420
tagccgctat cgcaaacaga tggcgggtgaa aaagtatctg gcggcgggtg tgggctaata 480
agaccaaattg tcg 493

```

<210> 27

<211> 493

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述：合成 DNA

<400> 27

```

gatccgacat ttggtcttat tagcccagca ccgcccgcag atactttttc accgccatct 60
gtttgcgata goggtatag ctatcggtaa agatgccatc gctatgccgg ccgtcaatgc 120
cgctgctgtg gtggtggtgg tgatggctag ccataatggc agaaccagaa ccggccaggt 180
tagcgtcgag gaactctttc aactgacctt tagacagtgc acccaatttg gttgccgcca 240
cttcacogtt tttgaacagc agcagagtcg ggataccacg gatgccatat ttccggcgag 300
tgccagggtt ttgatcgatg ttcagttttg caacggtcag tttgccctga tattcgtcag 360
cgatttcacg cagaatcggg gogatcattt tgcacggacc gcaccactct gccagaaat 420
cgacgaggat cggcccgtcc gctttgagta catccgtgtc aaaactgtcg tcagtcaggt 480
ggaccaatgg tcg 493

```

<210> 28

<211> 234

<212> DNA

<213> 人

<220>

<221> CDS

<222> (1).. (234)

<400> 28

```

acc cct gcc tac cta agc cag gcc aca gag ctc atc acc cag gcc ctg 48
Thr Pro Ala Tyr Leu Ser Gln Ala Thr Glu Leu Ile Thr Gln Ala Leu
1 5 10 15

cgg gat gag aag gca ggc gct tac gct gct gca ctc cag ggc tat cga 96
Arg Asp Glu Lys Ala Gly Ala Tyr Ala Ala Ala Leu Gln Gly Tyr Arg
20 25 30

```

gac ggc gtg cac gtc ttg ctt cag gga gtc ccc agt gac ccg ttg cct 144
 Asp Gly Val His Val Leu Leu Gln Gly Val Pro Ser Asp Pro Leu Pro
 35 40 45

gcc cgc cag gaa ggt gtg aag aag aag gca gct gag tac ctg aag cgg 192
 Ala Arg Gln Glu Gly Val Lys Lys Lys Ala Ala Glu Tyr Leu Lys Arg
 50 55 60

gca gag gag atc ctg cgc ctg cac ctg tct caa ctc cca ccc 234
 Ala Glu Glu Ile Leu Arg Leu His Leu Ser Gln Leu Pro Pro
 65 70 75

<210> 29
 <211> 78
 <212> PRT
 <213> 人

<400> 29
 Thr Pro Ala Tyr Leu Ser Gln Ala Thr Glu Leu Ile Thr Gln Ala Leu
 1 5 10 15

Arg Asp Glu Lys Ala Gly Ala Tyr Ala Ala Ala Leu Gln Gly Tyr Arg
 20 25 30

Asp Gly Val His Val Leu Leu Gln Gly Val Pro Ser Asp Pro Leu Pro
 35 40 45

Ala Arg Gln Glu Gly Val Lys Lys Lys Ala Ala Glu Tyr Leu Lys Arg
 50 55 60

Ala Glu Glu Ile Leu Arg Leu His Leu Ser Gln Leu Pro Pro
 65 70 75

<210> 30
 <211> 210
 <212> DNA
 <213> 小鼠 (Mus musculus)

<220>
 <221> CDS
 <222> (1).. (210)

<400> 30
 tac ctt ggc cag gcc aca gag ctc atc act caa gcc ctg cgg aat gaa 48

```

Tyr Leu Gly Gln Ala Thr Glu Leu Ile Thr Gln Ala Leu Arg Asn Glu
 1             5             10             15
aag gct ggc gcc tac gct gct gct ctc caa ggc tac caa gat ggg gtg 96
Lys Ala Gly Ala Tyr Ala Ala Ala Leu Gln Gly Tyr Gln Asp Gly Val
      20             25             30
cac att ctt ctg cag ggt gtc tca ggt gac cca tcc cct gcc cgc cgg 144
His Ile Leu Leu Gln Gly Val Ser Gly Asp Pro Ser Pro Ala Arg Arg
      35             40             45
gaa ggt gtg aag aag aag gca gcc gag tac ctg aag cgg gca gag atg 192
Glu Gly Val Lys Lys Lys Ala Ala Glu Tyr Leu Lys Arg Ala Glu Met
      50             55             60
ctt cac aca cac ttg ccc 210
Leu His Thr His Leu Pro
65             70

```

<210> 31
 <211> 70
 <212> PRT
 <213> 小鼠 (*Mus musculus*)

```

<400> 31
Tyr Leu Gly Gln Ala Thr Glu Leu Ile Thr Gln Ala Leu Arg Asn Glu
 1             5             10             15
Lys Ala Gly Ala Tyr Ala Ala Ala Leu Gln Gly Tyr Gln Asp Gly Val
      20             25             30
His Ile Leu Leu Gln Gly Val Ser Gly Asp Pro Ser Pro Ala Arg Arg
      35             40             45
Glu Gly Val Lys Lys Lys Ala Ala Glu Tyr Leu Lys Arg Ala Glu Met
      50             55             60
Leu His Thr His Leu Pro
65             70

```

<210> 32
 <211> 212
 <212> DNA
 <213> 酿酒酵母 (*Saccharomyces cerevisiae*)

<220>

<221> CDS

<222> (1).. (210)

<400> 32

atg agc acg gga gat ttt tta act aag gga ata gaa ctg gtt caa aag 48
 Met Ser Thr Gly Asp Phe Leu Thr Lys Gly Ile Glu Leu Val Gln Lys
 1 5 10 15

gcc att gac ctg gac aca gcc acg cag tat gag gag gcg tat aca tca 96
 Ala Ile Asp Leu Asp Thr Ala Thr Gln Tyr Glu Glu Ala Tyr Thr Ser
 20 25 30

tac tat aac gga tta gat tat ttg atg ttg gct ttg aaa tac gag aaa 144
 Tyr Tyr Asn Gly Leu Asp Tyr Leu Met Leu Ala Leu Lys Tyr Glu Lys
 35 40 45

aac cct aag tcg aag gat ttg ata aga gcg aag ttt aca gag tat cta 192
 Asn Pro Lys Ser Lys Asp Leu Ile Arg Ala Lys Phe Thr Glu Tyr Leu
 50 55 60

aac cgt gca gag cag ttg aa 212
 Asn Arg Ala Glu Gln Leu
 65 70

<210> 33

<211> 70

<212> PRT

<213> 酿酒酵母 (*Saccharomyces cerevisiae*)

<400> 33

Met Ser Thr Gly Asp Phe Leu Thr Lys Gly Ile Glu Leu Val Gln Lys
 1 5 10 15

Ala Ile Asp Leu Asp Thr Ala Thr Gln Tyr Glu Glu Ala Tyr Thr Ser
 20 25 30

Tyr Tyr Asn Gly Leu Asp Tyr Leu Met Leu Ala Leu Lys Tyr Glu Lys
 35 40 45

Asn Pro Lys Ser Lys Asp Leu Ile Arg Ala Lys Phe Thr Glu Tyr Leu
 50 55 60

Asn Arg Ala Glu Gln Leu
 65 70

<210> 34
 <211> 231
 <212> DNA
 <213> 人

<220>
 <221> CDS
 <222> (1).. (231)

<400> 34
 atg tca tcc act tcg ccc aac ctc cag aaa gcg ata gat ctg gct agc 48
 Met Ser Ser Thr Ser Pro Asn Leu Gln Lys Ala Ile Asp Leu Ala Ser
 1 5 10 15
 aaa gca gcg caa gaa gac aag gct ggg aac tac gaa gaa gcc ctt cag 96
 Lys Ala Ala Gln Glu Asp Lys Ala Gly Asn Tyr Glu Glu Ala Leu Gln
 20 25 30
 ctc tat cag cat gct gtg cag tat ttt ctt cat gtc gtt aaa tat gaa 144
 Leu Tyr Gln His Ala Val Gln Tyr Phe Leu His Val Val Lys Tyr Glu
 35 40 45
 gca cag ggt gat aaa gcc aag caa agt atc agg gca aag tgt aca gaa 192
 Ala Gln Gly Asp Lys Ala Lys Gln Ser Ile Arg Ala Lys Cys Thr Glu
 50 55 60
 tat ctt gat aga gca gaa aaa cta aag gag tac ctg aaa 231
 Tyr Leu Asp Arg Ala Glu Lys Leu Lys Glu Tyr Leu Lys
 65 70 75

<210> 35
 <211> 77
 <212> PRT
 <213> 人

<400> 35
 Met Ser Ser Thr Ser Pro Asn Leu Gln Lys Ala Ile Asp Leu Ala Ser
 1 5 10 15
 Lys Ala Ala Gln Glu Asp Lys Ala Gly Asn Tyr Glu Glu Ala Leu Gln
 20 25 30
 Leu Tyr Gln His Ala Val Gln Tyr Phe Leu His Val Val Lys Tyr Glu
 35 40 45
 Ala Gln Gly Asp Lys Ala Lys Gln Ser Ile Arg Ala Lys Cys Thr Glu

50 55 60
 Tyr Leu Asp Arg Ala Glu Lys Leu Lys Glu Tyr Leu Lys
 65 70 75

<210> 36
 <211> 231
 <212> DNA
 <213> 小鼠 (Mus musculus)

<220>
 <221> CDS
 <222> (1).. (231)

<400> 36
 atg gcg tcc acg aac acc aac ctg cag aaa gca ata gat ctt gca agc 48
 Met Ala Ser Thr Asn Thr Asn Leu Gln Lys Ala Ile Asp Leu Ala Ser
 1 5 10 15
 aaa gca gcc cag gaa gac aaa gct ggg aac tat gag gaa gct ctt caa 96
 Lys Ala Ala Gln Glu Asp Lys Ala Gly Asn Tyr Glu Glu Ala Leu Gln
 20 25 30
 ctc tac caa cat gct gtg cag tat ttt ctc cat gtt gtt aaa tat gaa 144
 Leu Tyr Gln His Ala Val Gln Tyr Phe Leu His Val Val Lys Tyr Glu
 35 40 45
 gca caa ggt gat aaa gcc aag caa agt atc agg gcc aag tgt aca gaa 192
 Ala Gln Gly Asp Lys Ala Lys Gln Ser Ile Arg Ala Lys Cys Thr Glu
 50 55 60
 tat ctc gat aga gca gaa aaa cta aag gaa tat ctg aag 231
 Tyr Leu Asp Arg Ala Glu Lys Leu Lys Glu Tyr Leu Lys
 65 70 75

<210> 37
 <211> 77
 <212> PRT
 <213> 小鼠 (Mus musculus)

<400> 37
 Met Ala Ser Thr Asn Thr Asn Leu Gln Lys Ala Ile Asp Leu Ala Ser
 1 5 10 15
 Lys Ala Ala Gln Glu Asp Lys Ala Gly Asn Tyr Glu Glu Ala Leu Gln

	20		25		30										
Leu	Tyr	Gln	His	Ala	Val	Gln	Tyr	Phe	Leu	His	Val	Val	Lys	Tyr	Glu
	35						40						45		
Ala	Gln	Gly	Asp	Lys	Ala	Lys	Gln	Ser	Ile	Arg	Ala	Lys	Cys	Thr	Glu
	50					55					60				
Tyr	Leu	Asp	Arg	Ala	Glu	Lys	Leu	Lys	Glu	Tyr	Leu	Lys			
	65				70				75						

<210> 38
 <211> 21
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述 : 合成 DNA

<400> 38
 acccctgcct acctaagcca g 21

<210> 39
 <211> 22
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述 : 合成 DNA

<400> 39
 ggattacagg tgcaggcgca gg 22

<210> 40
 <211> 19
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述 : 合成 DNA

<400> 40
 taccttgcc aggccacag 19

<210> 41
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 41
ggattagggc aagtgtgtgt gaagc 25

<210> 42
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 42
atgagcacgg gagatTTTTT aactaag 27

<210> 43
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 43
ggattattcc aagtgtttt tcaactgctc 30

<210> 44
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述：合成 DNA

<400> 44
atgtcatcca cttcgccc 18

<210> 45
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述 : 合成 DNA

<400> 45
ggattatttc aggtactcct ttagtttttc tgctc 35

<210> 46
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述 : 合成 DNA

<400> 46
atggcgtcca cgaacacc 18

<210> 47
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述 : 合成 DNA

<400> 47
ggattacttc agatattcct ttagtttttc tgctctatc 39

<210> 48
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述 : 合成 DNA

<400> 48
agggacgaga cggccc 16

<210> 49
 <211> 28
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述：合成 DNA

<400> 49
 ggattaaggt gaagagtaca ggctgctg 28

<210> 50
 <211> 29
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述：合成 DNA

<400> 50
 atgattaagg tgaagagtac aggctgctg 29

<210> 51
 <211> 190
 <212> DNA
 <213> 人

<220>
 <221> CDS
 <222> (1)..(189)

<400> 51
 aag agg gac gag acg gcc ctg cag gag gag gag gag ctg cag ctg gcc 48
 Lys Arg Asp Glu Thr Ala Leu Gln Glu Glu Glu Glu Leu Gln Leu Ala
 1 5 10 15
 ctg gcg ctg tca cag tca gag gcg gag gag aag gag agg ctg aga cag 96
 Leu Ala Leu Ser Gln Ser Glu Ala Glu Glu Lys Glu Arg Leu Arg Gln
 20 25 30
 aag tcc acg tac act tcg tac ccc aag gcg gag ccc atg ccc tcg gcc 144
 Lys Ser Thr Tyr Thr Ser Tyr Pro Lys Ala Glu Pro Met Pro Ser Ala
 35 40 45
 tcc tca gcg ccc ccc gcc agc agc ctg tac tct tca cct gtg aac t 190

Ser Ser Ala Pro Pro Ala Ser Ser Leu Tyr Ser Ser Pro Val Asn
 50 55 60

<210> 52

<211> 63

<212> PRT

<213> 人

<400> 52

Lys Arg Asp Glu Thr Ala Leu Gln Glu Glu Glu Leu Gln Leu Ala
 1 5 10 15

Leu Ala Leu Ser Gln Ser Glu Ala Glu Glu Lys Glu Arg Leu Arg Gln
 20 25 30

Lys Ser Thr Tyr Thr Ser Tyr Pro Lys Ala Glu Pro Met Pro Ser Ala
 35 40 45

Ser Ser Ala Pro Pro Ala Ser Ser Leu Tyr Ser Ser Pro Val Asn
 50 55 60

<210> 53

<211> 10

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 53

ggnnnnntta

10

<210> 54

<211> 10

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 54

atgnnnnntta

10

<210> 55

<211> 465
 <212> DNA
 <213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 55

```
cacctgactg acgacagttt tgacacggat gtactcaaag cggacggggc gatcctcgtc 60
gatttctggg cagagtgggt cggtcctgtc aaaatgatcg ccccgattct ggatgaaatc 120
gctgacgaat atcagggcaa actgaccgtt gcaaaactga acatcgatca aaaccctggc 180
actgcgccga aatatggcat cgttggtatc cggactctgc tgctgttcaa aaacggtgaa 240
gtggcggcaa ccaaagtggg tgcactgtct aaagggtcagt tgaaagagtt cctcgacgct 300
aacctggccg gttctggttc tggccatatg gctagccatc accaccacca ccacagcagc 360
ggcattgacg gccggcatag cgatggcatc tttaaccgata gctatagccg ctatcgcaaa 420
cagatggcgg tgaaaaagta tctggcggcg gtgctgggct aataa 465
```

<210> 56
 <211> 465
 <212> DNA
 <213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 56

```
ttattagccc agcaccgccg ccagatactt tttcaccgcc atctgtttgc gatagcggct 60
atagctatcg gtaaagatgc catcgctatg ccggccgtca atgccgctgc tgtgttggtg 120
gtgtgatggt ctgacatgat ggccagaacc agaaccggcc aggttagcgt cgaggaaact 180
tttcaactga ccttagaca gtgcaccac tttggttgcc gccacttcac cgtttttgaa 240
cagcagcaga gtcgggatac cacggatgcc atatttcggc gcagtgccag ggttttgatc 300
gatgttcagt tttgcaacgg tcagtttgcc ctgatattcg tcagcgattt catccagaat 360
cggggcgatc attttgacg gaccgcacca ctctgccag aaatcgacga ggatcgcccc 420
gtccgctttg agtacatccg tgtcaaaact gtcgtcagtc aggtg 465
```

<210> 57
 <211> 4974
 <212> DNA
 <213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 57

```
ggatccccga attcccggtt cgactcgagc ggccgcatcg tgactgactg acgatctgcc 60
tcgcgcgttt cggatgatgc ggtgaaaacc tctgacacat gcagctcccg gagacggtca 120
```

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 180
 ttggcgggtg tcggggcgca gccatgaccc agtcacgtag cgatagcgga gtgtataatt 240
 cttgaagaog aaagggcctc gtgatacgcc tttttttata ggtaaatgtc atgataataa 300
 tggtttctta gacgtcaggt ggcacttttc ggggaaatgt gcgcggaacc cctatttgtt 360
 tttttttcta aatacattca aatatgtatc cgtcatgag acaataaccc tgataaatgc 420
 ttcaataata ttgaaaaagg aagagtatga gtattcaaca ttccgtgtc gcccttattc 480
 ccttttttgc ggcattttgc ctctctgttt ttgctcacc agaaacgctg gtgaaagtaa 540
 aagatgctga agatcagttg ggtgcacgag tgggttacat cgaactggat ctcaacagcg 600
 gtaagatcct tgagagtttt cgccccgaag aacgttttcc aatgatgagc acttttaag 660
 ttctgtctatg tggcgcggtt ttatccgtg ttgacgccg gcaagagcaa ctcggtcgcc 720
 gcatacacta ttctcagaat gacttggttg agtactcacc agtcacagaa aagcatctta 780
 cggatggcat gacagtaaga gaattatgca gtgctgcat aaccatgagt gataacactg 840
 cggccaactt acttctgaca acgacggag gaccgaagga gctaaccgct ttttgcaca 900
 acatggggga tcatgtaact cgccttgatc gttgggaacc ggagctgaat gaagccatac 960
 caaacgacga gcgtgacacc acgatgocct cagcaatggc aacaacgtt cgcaactat 1020
 taactggcga actacttact ctacttccc ggcaacaatt aatagactgg atggaggcgg 1080
 ataaagtgc aggaccactt ctgcgtcgg cccttccggc tggctggttt attgctgata 1140
 aatctggagc cggtagcgt gggctcgcg gtatcattgc agcactggg ccagatggta 1200
 agccctcccg tatcgtagtt atctacacca cggggagcca ggcaactat gatgaacgaa 1260
 atagacagat cgtgagata ggtgcctcac tgattaagca ttgtaactg tcagaccaag 1320
 ttactcata tatactttag attgatitaa aacttcattt ttaattitaa aggatctagg 1380
 tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact 1440
 gagcgtcaga ccccgtagaa aagatcaaag gatcttctt agatcccttt tttctgcgcg 1500
 taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggtc 1560
 aagagctacc aactctttt ccgaaggtaa ctggcttcag cagagcgag ataccaaata 1620
 ctgtccttct agttagccg tagttaggcc accacttcaa gaactctgta gcaccgccta 1680
 catacctgc tctgctaact ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc 1740
 ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg 1800
 ggggttcgtg cacacagccc agcttggagc gaacgacct caccgaactg agatacctac 1860
 agcgtgagct atgagaaagc gccacgcttc ccgaaggag aaaggcggac aggtatccgg 1920
 taagcggcag ggtcggaaca ggagagcgca cgaggagct tccaggggga aacgcctggt 1980
 atctttatag tctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgt 2040
 cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggccttttta cggttcctgg 2100
 ccttttctg gccttttct cactgttct ttcctgcgtt atccctgat tctgtggata 2160
 accgtattac cgcctttgag tgagctgata ccgtcgcgg cagccgaacg accgagcgca 2220
 gcgagtcagt gagcgaggaa gcggaagagc gcctgatcg gtattttct cttacgcato 2280
 tgtcggtat ttacacccgc ataaattccg acaccatcga atggtgcaaa acctttcgcg 2340
 gtatggcatg atagcggcga gaagagagtc aattcagggt ggtgaatgt aaaccagtaa 2400
 cgttatacga tgtcgagag tatgcgggt tctcttatca gaccgtttcc cgcgtggtga 2460
 accaggccag ccacgtttct gcgaaaacgc gggaaaaagt ggaagcgcg atggcgagc 2520
 tgaattacat tcccaaccgc gtggcacaac aactggcggg caaacagtcg ttgctgattg 2580
 gcgttgccac ctccagtctg gccctgcacg cgcgtcgca aattgtcgcg gcgattaaat 2640
 ctgcgcccga tcaactgggt gccagcgtgg tgggtcgat ggtagaagca agcggcgctc 2700
 aagcctgtaa agcggcgggt cacaatcttc tcgcgcaac cgtcagtggt ctgatcatta 2760
 actatccgct ggatgaccag gatgccattg ctgtggaagc tgcctgcact aatgttccgg 2820
 cgttatttct tgatgtctct gaccagacac ccatcaacag tattatttcc tccatgaag 2880
 acggtacgcg actggcggtg gagcatctgg tcgcattggg tcaccagcaa atcgcgctgt 2940
 tagcggggcc attaagttct gtctcggcgc gtctgcgtct ggctggctgg cataaatatc 3000

```

tcactcgcaa tcaaattcag ccgatagcgg aacgggaagg cgactggagt gccatgtccg 3060
gttttcaaca aacctgcaa atgctgaatg agggcatcgt tcccactgcg atgctggttg 3120
ccaacgatca gatggcgtg ggcgcaatgc gcgccattac cgagtccggg ctgcgcgttg 3180
gtgcggatat ctcggtagt ggatacgacg ataccgaaga cagctcatgt tataccccgc 3240
cgtttaaccac catcaaacag gattttcgcc tgcctgggca aaccagcgtg gaccgcttgc 3300
tgcaactctc tcagggccag gcggtgaagg gcaatcagct gttgcccgtc tcaactgtga 3360
aaagaaaaac caccctggcg cccaatacgc aaaccgcctc tcccgcgcgc ttggccgatt 3420
cattaatgca gctggcacga caggtttccc gactggaaag cgggcagtga gcgcaacgca 3480
attaatgtga gttagctcac tcattaggca cccaggcctt tacactttat gcttccggct 3540
cgtatgttgt gtggaattgt gagcggataa caatttcaca caggaaacag ctatgaccat 3600
gattacggat tcaactggcg tcgttttaca acgtcgtgac tgggaaaacc ctggcggttac 3660
ccaacttaat cgccttgca ccatccccc ttccgccagc tggcgtaata gcgaagaggc 3720
ccgcaccgat cgccttcccc aacagttgcg cagcctgaat ggcaatggc gctttgcctg 3780
gtttccggca ccagaagcgg tgcggaaaag ctggctggag tgcgatcttc ctgaggccga 3840
tactgtctgc gtccctcaa actggcagat gcacggttac gatgogccca tctacaccaa 3900
cgtaacctat ccattacgg tcaatccgcc gttgttccc acggagaatc cgacgggttg 3960
ttactcgtc acatttaatg ttgatgaaag ctggctacag gaaggccaga cgcgaattat 4020
ttttgatggc gttggaatta cgttatcgac tgcacggtgc accaatgctt ctggcgtcag 4080
gcagccatcg gaagctgttg tatggctgtg caggctgtaa atcactgcat aattcgtgtc 4140
gctcaaggcg cactcccggt ctggataatg ttttttgcgc cgacatcata acggttcttg 4200
caaataattc gaaatgagct gttgacaatt aatcatcggc tcgtataatg tgtggaattg 4260
tgagcggata acaatttcac acaggaaaca gtattcatgt cccctatact aggttatttg 4320
aaaattaagg gccttgtgca acccactcga ctcttttg aatatcttga agaaaaatat 4380
gaagagcatt tgtatgagcg cgatgaagg gataaatggc gaaacaaaaa gtttgaattg 4440
ggtttggagt ttccaatct tccttattat attgatggtg atgttaaatt aacacagctc 4500
atggccatca tacgttatat agctgacaag cacaacatgt tgggtggttg tccaaaagag 4560
cgtgcagaga tttcaatgct tgaaggagcg gttttggata ttagatacgg tgtttcgaga 4620
attgcatata gtaaagactt tgaactctc aaagttgatt ttcttagcaa gctacctgaa 4680
atgtgaaaaa tgttcgaaga tcgtttatgt cataaaacat atttaaatgg tgatcatgta 4740
accatcctg acttcatgtt gtatgacgct ctgatgttg tttatacat ggacccaatg 4800
tgcttgatg cgttccaaa attagtttgt tttaaaaaac gtattgaagc tatcccacaa 4860
attgataagt acttgaaac cagcaagtat atagcatggc ctttgacagg ctggcaagcc 4920
acgtttgtg gtggcgacca tcctccaaaa tcggatctgg ttccgcgtgg atcc 4974

```

<210> 58
 <211> 4974
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述：合成 DNA

<400> 58
 ggatccacgc ggaaccagat ccgattttgg aggatggtcg ccaccaccaa acgtggcttg 60
 ccagccctgc aaaggccatg ctatatactt gctggatttc aagtacttat caatttgttg 120
 gatagcttca atacgttttt taaaacaaac taattttggg aacgcataca ggcacatttg 180
 gtccatgtat aaaacaacat caagagcgtc atacaacatg aagtcaggat gggttacatg 240

atcaccattt aaatatgttt tatgacataa acgatcttcg aacattttca gcatttcagg 300
 tagcttgcta agaaaatcaa ctttgagagt ttcaaagtct ttactatatg caattctcga 360
 aacaccgtat ctaatatcca aaaccgctcc ttcaagcatt gaaatctctg cacgctcttt 420
 tggacaacca cccaacatgt tgtgcttgtc agctatataa cgtatgatgg ccatagactg 480
 tgttaattta acatcaccat caatataata aggaagattg ggaaactcca aaccaattc 540
 aaactttttg tticgccatt tatcaccttc atcgcgctca tacaatgct cttcatattt 600
 ttcttcaaga tattccaaaa gaagtcgagt gggttgaca aggccttaaa tttccaata 660
 acctagtata ggggacatga atactgtttc ctgtgtgaaa ttgttatccg ctcacaattc 720
 cacacattat acgagccgat gattaattgt caacagctca ttccagaata tttgccagaa 780
 ccgttatgat gtcggcgcaa aaaacattat ccagaacggg agtgcgcctt gagcgacacg 840
 aattatgcag tgatttacga cctgcacagc cataccacag cttccgatgg ctgcctgacg 900
 ccagaagcat tggcgaccg tgcagtcgat aacgtaattc caacgccatc aaaaaataatt 960
 cgcgtctggc cttcctgtag ccagctttca tcaacattaa atgtgagcga gtaacaaccc 1020
 gtccgattct ccgtgggaac aaacggcgga ttgaccgtaa tgggataggt tacgttgggtg 1080
 tagatgggag catcgtaacc gtgcatctgc cagtttgagg ggacgacgac agtatcggcc 1140
 tcaggaagat cgcactccag ccagctttcc ggcaaccgct ctggtgccgg aaaccaggca 1200
 aagcgccatt cgcattccag gctgcgcaac tgttggaag ggcgatcgtt gcgggcctct 1260
 tcgctattac gccagctggc gaaaggggga tgtctgcaa ggcgattaag ttgggtaacg 1320
 ccagggtttt ccagtcacg acgttgtaaa acgacggcca gtgaatccgt aatcatggtc 1380
 atagctgttt cctgtgtgaa attgttatcc gtcacaatt ccacacaaca tacgagccgg 1440
 aagcataaag tgtaagcct ggggtgccta atgagtgagc taactcacat taattgcgtt 1500
 gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt aatgaatcgg 1560
 ccaacgcgag gggagaggcg gtttgcgtat tgggcgccag ggtggtttt cttttacca 1620
 gtgagacggg caacagctga ttgcccttca ccgcctggcc ctgagagagt tgcagcaagc 1680
 ggtccacgct ggtttgcccc agcaggcgaa aatcctgttt gatggtggtt aacggcggga 1740
 tataacatga gctgtcttcg gtatcgtcgt atcccactac cgagatatcc gcaccaacgc 1800
 gcagcccggg ctcggtaatg gcgcgcattg cgcacagcgc catctgatcg ttggcaacca 1860
 gcacgcgagt gggaaacgat ccctcattca gcatttgcat ggtttgttga aaaccggaca 1920
 tggcactcca gtcgccttcc cgttcgcgta tcggctgaat ttgattgcga gtgagatatt 1980
 tatgccagcc agccagacgc agacgcgcg agacagaact taatgggccc gctaacagcg 2040
 cgatttgctg gtgacccaat gcgaccagat gtcacacgcc cagtcgcgta ccgtcttcat 2100
 gggagaaaat aatactgttg atgggtgtct ggtcagagac atcaagaaat aacgccggaa 2160
 cattagtga ggcagcttcc acagcaatgg catcctggtc atccagcgga tagttaatga 2220
 tcagccact gacgcgttgc gcgagaagat tgtgcaccgc cgctttacag gcttcgacgc 2280
 cgcttogttc taccatcgac accaccacgc tggcaccag ttgatcgcg cgagatttaa 2340
 tcgcccgcac aatttgcgac ggcgcgtgca gggccagact ggaggtggca acgccaatca 2400
 gcaacgactg tttccccgcc agttgttggt ccacgcggtt gggaatgtaa ttcagctccg 2460
 ccacgcgcgc ttcactttt tccgcggtt tgcagaaac gtggctggcc tggttacca 2520
 cgcgggaaac ggtotgataa gagacaccgg catactctgc gacatcgtat aacgttactg 2580
 gtttcacatt caccaccctg aattgactct cttccgggog ctatcatgcc ataccgcgaa 2640
 aggttttgca ccattcgatg gtgtoggaat ttatgcggtg tgaataaccg cacagatcg 2700
 taaggagaaa ataccgcac aggcgctctt ccgcttctc gtcactgac tcgctgcgt 2760
 cggctcgttc gctgcggcga gcggtatcag ctactcaaa ggcggtataa cggttatcca 2820
 cagaatcagg ggataacgca ggaagaaca tgtgagcaaa aggcagcaa aaggccagga 2880
 accgtaaaaa ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc 2940
 acaaaaatcg acgtcaagt cagagggtggc gaaacccgac aggactataa agataccagg 3000
 cgtttcccc tggagctcc ctctgcgt ctctgttcc gaccctgcg cttaccgat 3060
 acctgtccgc cttctccct tcgggaagcg tggcgcttcc tcatagctca cgtgttaggt 3120

```

atctcagttc ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgttc 3180
agcccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacaog 3240
acttatcgcc actggcagca gccactggta acaggattag cagagcgagg tatgtaggog 3300
gtgctacaga gttcttgaag tgggtggccta actacggcta cactagaagg acagtatttg 3360
gtatctgcgc tctgctgaag ccagttacot tccgaaaaag agttggtagc tcttgatccg 3420
gcaaaacaaac caccgctggt agcgggtggt tttttgttg caagcagcag attacgcgca 3480
gaaaaaaagg atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga 3540
acgaaaaactc acgttaaggg attttggta tgagattatc aaaaaggatc ttcacctaga 3600
tccttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag taaacttggg 3660
ctgacagtta ccaatgctta atcagtggag cacctatctc agcagatctg ctatttcgtt 3720
catccatagt tgcctggctc cccgtgggtg agataactac gatacgggag ggcttaccat 3780
ctggccccag tctgcaatg ataaccgcag acccagctc accggctcca gatttatcag 3840
caataaacca gccagccgga agggccgagc gcagaagtgg tctgcaact ttatccgct 3900
ccatccagtc tattaattgt tgcggggaag ctagagtaag tagttcgcca gtaaatagtt 3960
tgcgcaacgt tgttgccatt gctgcaggca tctgtgtgac acgctcgtc ttggtatgg 4020
cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatcccc atgttgtgca 4080
aaaaagcggg tagctcttcc ggtcctcoga tctgtgtcag aagtaagtig gccgcagtgt 4140
tatcactcat ggttatggca gcactgcata attctcttac tgtcatgcca tccgtaagat 4200
gcttttctgt gactgggtag tactcaacca agtcattctg agaatagtgt atgcggcgac 4260
cgagttgctc ttgcccgcg tcaacacggg ataataccgc gccacatagc agaacttta 4320
aagtgtcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt 4380
tgagatccag ttcgatgtaa cccactcgtg caccacactg atcttcagca tcttttactt 4440
tcaccagcgt tcttgggtga gcaaaaacag gaaggcaaaa tgcggcaaaa aagggaataa 4500
ggcgacacag gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt 4560
atcagggtta ttgtctcatg agcggatata tatttgaatg tatttagaaa aataaaciaa 4620
taggggttcc ggcacattt ccccgaaaag tgccacctga cgtctaagaa accattatta 4680
tcatgacatt aaactataaa aataggcgta tcacagggcc ctttcgtctt caagaattat 4740
acactccgct atcgtacgt gactgggtca tggctgcgcc ccgacacccg ccaacacccg 4800
ctgacgcgcc ctgacgggct tgtctgctcc cggcatccgc ttacagacaa gctgtgaccg 4860
tctccgggag ctgcatgtgt cagaggtttt caccgtcacc accgaaacgc gcgaggcaga 4920
tcgtcagtca gtcacgatgc ggccgctoga gtcgacccgg gaattcgggg atcc 4974

```

<210> 59

<211> 487

<212> DNA

<213> 人工序列

<220>

<223> 人工序列描述 : 合成 DNA

<400> 59

```

gaccattggt ccacctgact gacgacagtt ttgacacgga tgtactcaaa gcggacgggg 60
cgatcctcgt cgatttctgg gcagagtggg gcgggtccgtg caaaatgacg gccccgattc 120
tggatgaaat cgtgacgaa tatcagggca aactgaccgt tgcaaaactg aacatcgatc 180
aaaaccctgg cactgcgcg aaatatggca tccgtgggtat cccgactctg ctgctgttca 240
aaaacggtag agtggcggca accaaagtgg gtgactgtc taaaggtcag ttgaaagagt 300
tcctcgacgc taacctggcc ggttctggtt ctggccatat ggctagccat caccaccacc 360

```

```

accacagcag cggcattgac ggccggcata gcgatggcat ctttaccgat agctatagcc 420
gctatcgcaa acagatggcg gtgaaaaagt atctggcggc ggtgctgggc taataagacc 480
aaatgtc                                         487

```

<210> 60
 <211> 110
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述 : 合成 DNA

```

<400> 60
gaocggtcgt ctagcgatgg catctttacc gatagctata gccgctatcg caaacagatg 60
gcggtgaaaa agtatctggc ggcggtgctg ggctaataag accaatggtc          110

```

<210> 61
 <211> 259
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述 : 合成 DNA

```

<400> 61
gacgataagt ctcatatgca gattttcgtg aaaaccctga ccggcaaaac catcaccctg 60
gaagtggaac cgtccgacac catcgaaaac gttaaagcga aaatccagga caaagaaggc 120
atcccgccgg atcagcagcg tctgatcttc gcgggcaaac agctggaaga cggccgtacc 180
ctgtccgatt acaacatcca gaaagaatct accctgcacc tggctgctgcg tctgcgtggc 240
ggctaataga cttaaggtc                                         259

```

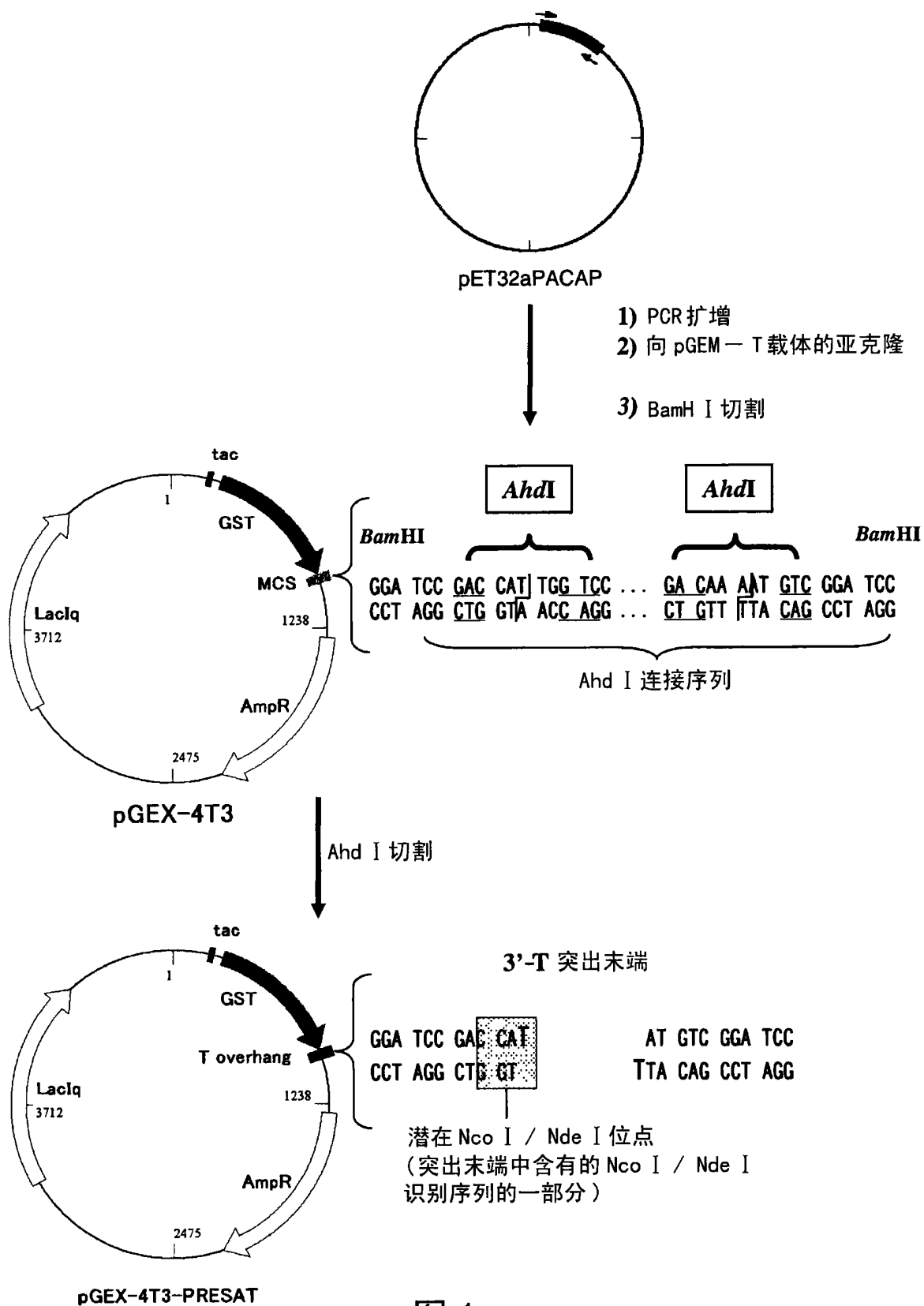


图 1

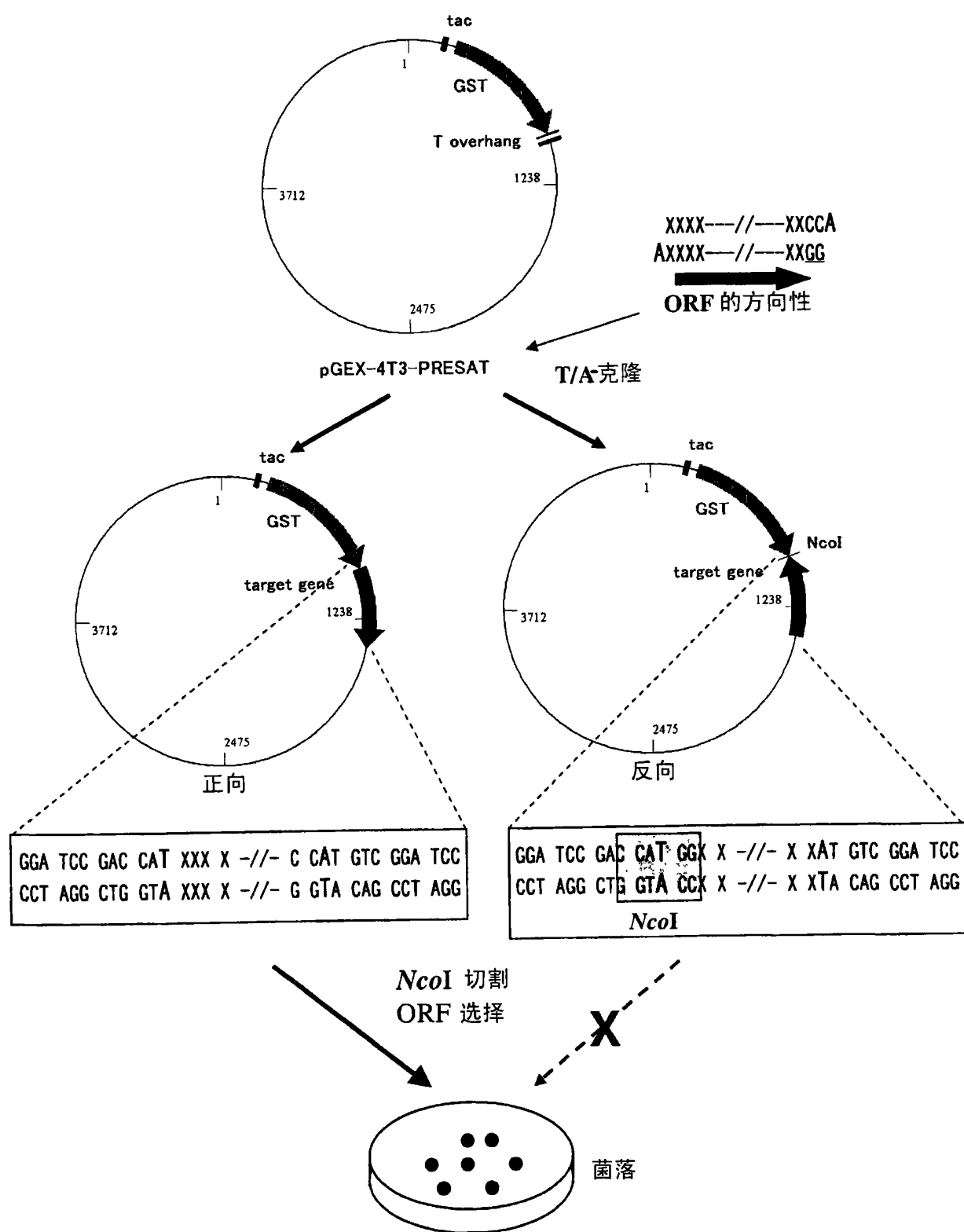


图 2

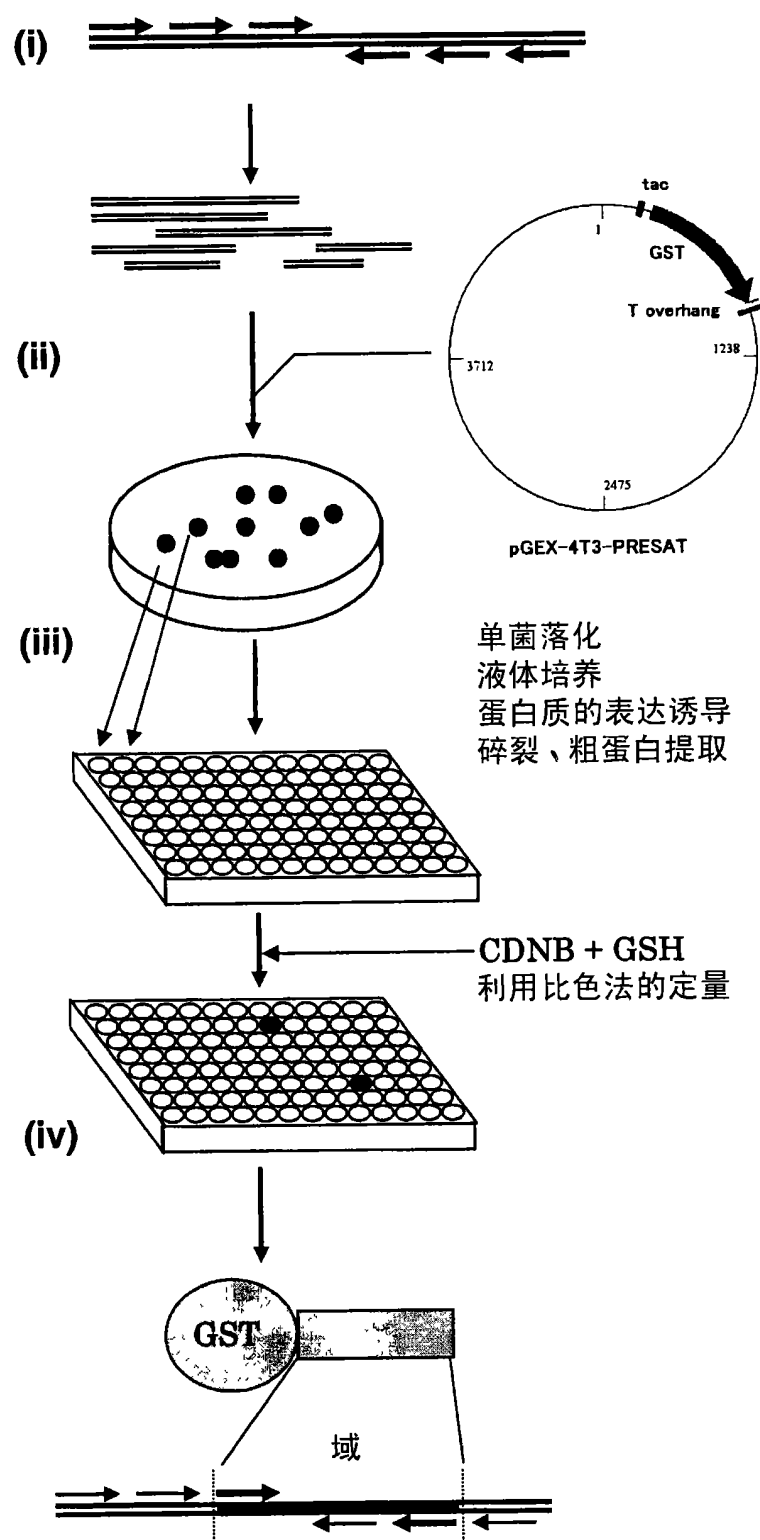


图 3

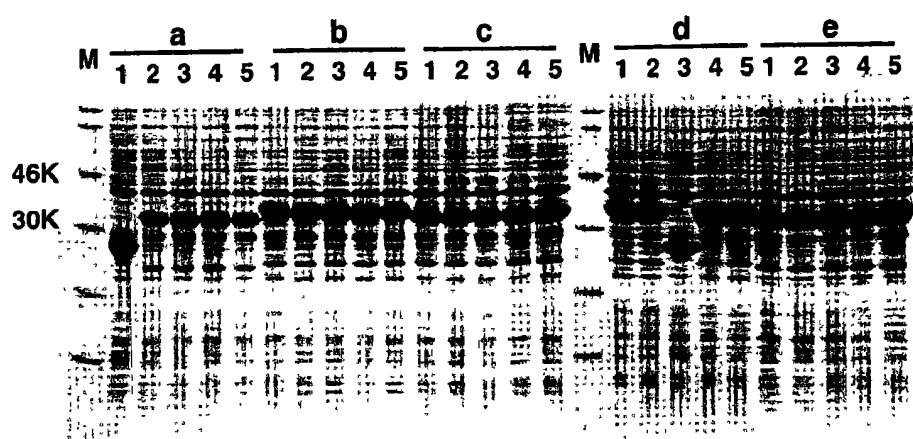


图 4

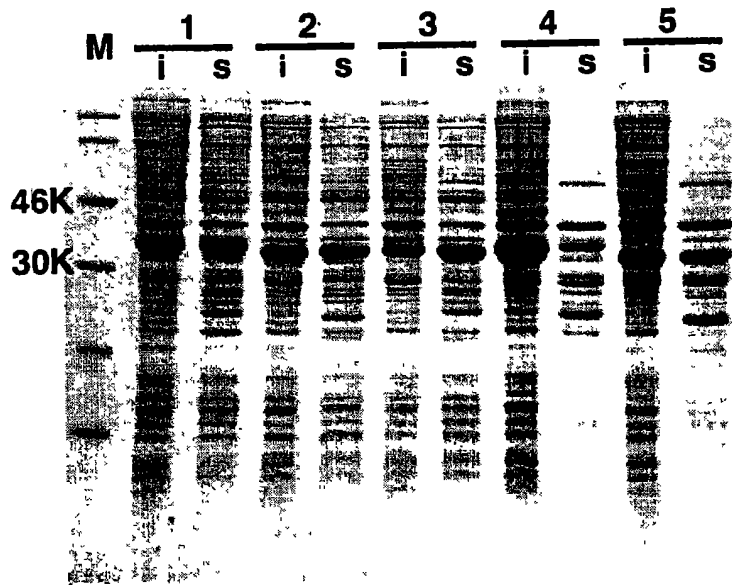


图 5

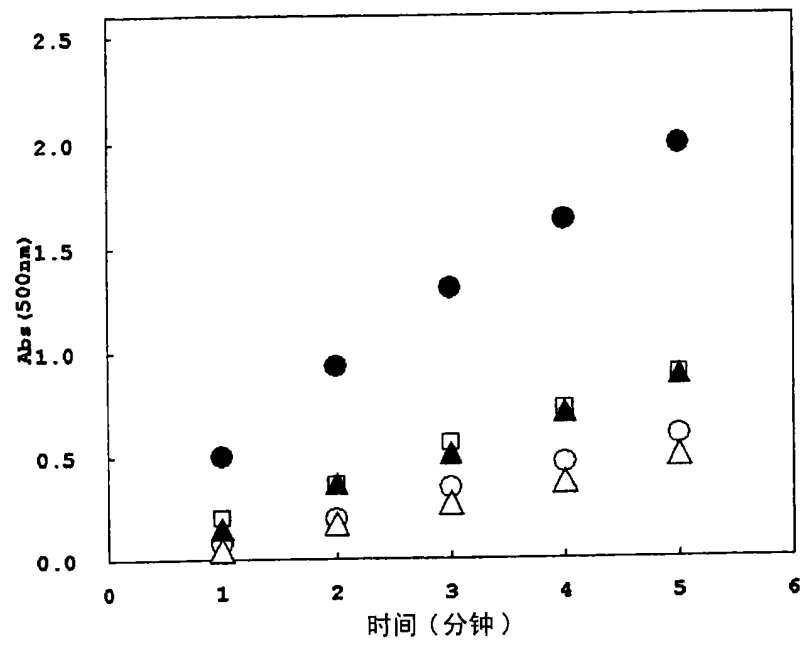


图 6

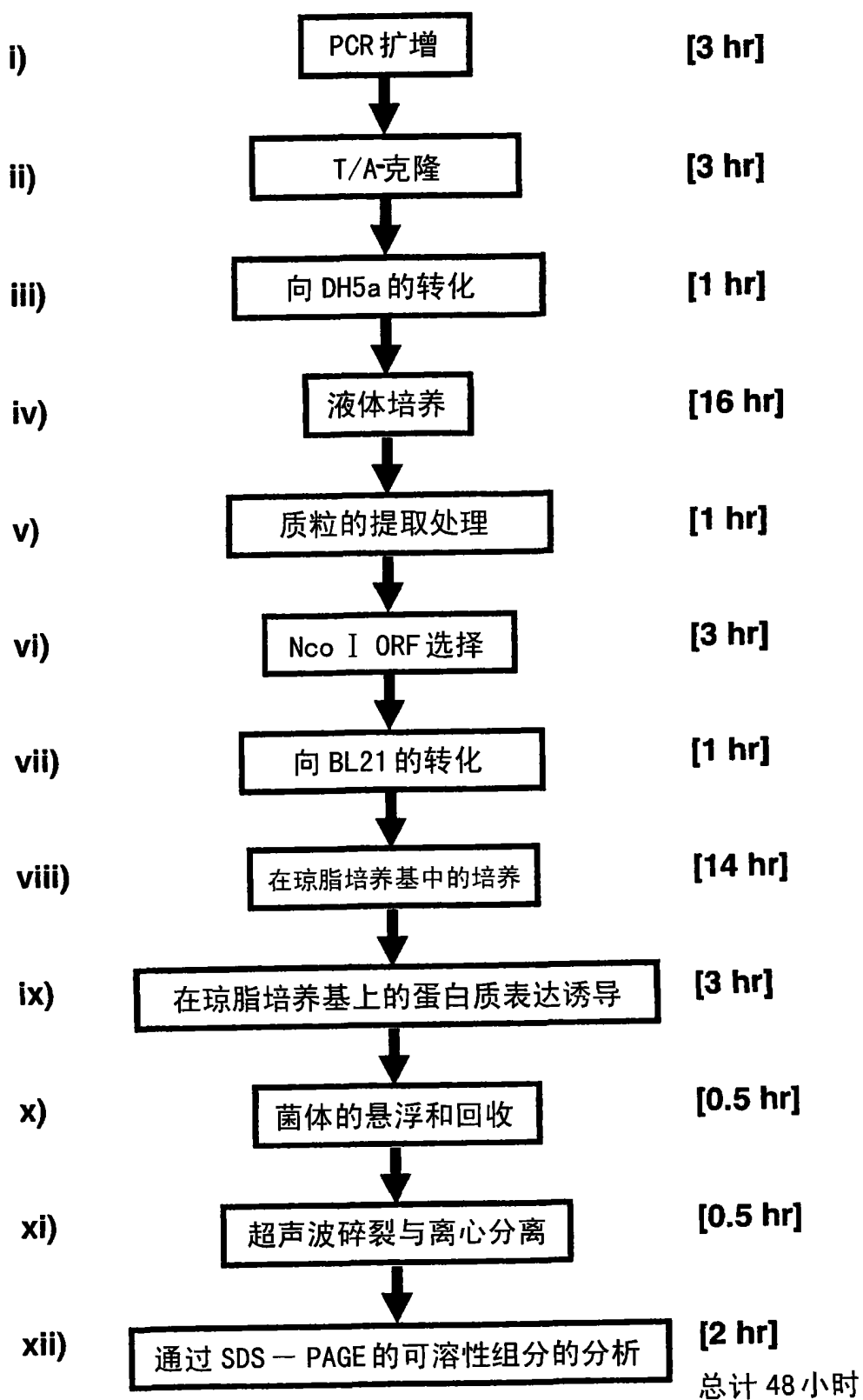


图 7

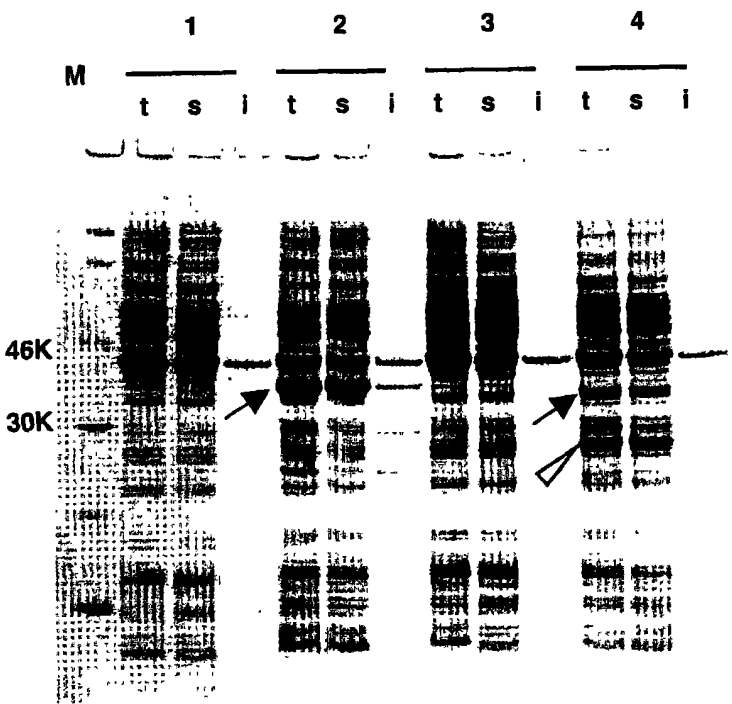


图 8