[19] 中华人民共和国国家知识产权局

[12] 发明专利说明书

[21] ZL 专利号 95193673.5

[51] Int. Cl[7]
C12N 15/38
C12N 1/21 C12N 5/10
C07K 14/05 A61K 39/245
A61P 31/22

[45] 授权公告日 2003 年 8 月 20 日

[11] 授权公告号 CN 1118573C

[22] 申请日 1995.4.13 [21] 申请号 95193673.5

[30] 优先权

[32] 1994.4.18 [33] US [31] 08/229,291

[86] 国际申请 PCT/US95/04611 1995.4.13

[87] 国际公布 WO95/28488 英 1995.10.26

[85] 进入国家阶段日期 1996.12.18

[71] 专利权人 阿维龙公司

地址 美国加利福尼亚州

[72] 发明人 R·施佩特 W·T·杰克曼

[56] 参考文献

JOURNAL OF GENERAL VIROLOGY, VOL. 73 1992 - 01 - 01 FINERTY ET AL " PROTECTIVE IMMUNIZATION AGAINST EPSTEIN - BARRVIRUS INDUCED

JOURNAL OF IMMUNOLOGICAL METHODS ,VOL. 108 1988 - 01 - 01 DAVID ET AL " EFFICIENT PURIFICATION OF EPSTEIN - BARR VIRUS MEMBTANE

THE EMBO JOURNAL ,OVL. 4, NOL. 12 1985 - 01 - 01 MACKETT ET AL," RECOMBINANT VACCINIA VIRUS INDUCES NEUTRALIZING ANTIBODIES IN RABBITS AGAIN ST EPSTEI

THE EMBO JOURNAL ,OVL. 4, NOL. 12 1985 - 01 - 01 MACKETT ET AL," RECOMBINANT VACCINIA VIRUS INDUCES NEUTRALIZING ANTIBODIES IN RABBITS AGAIN ST EPSTEI

审查员 周 霞

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 罗 宏 姜建成

权利要求书 2 页 说明书 42 页 附图 5 页

[54] 发明名称 gp350/220 的非剪接变异体

[57] 摘要

提供了含有 gp350 DNA 和氨基酸序列的组合物，以及含有所述序列的载体和宿主细胞。也提供了在没有 gp220 蛋白质产生的情况下用于重组制备 gp350 蛋白质的方法，提供了含有所述蛋白质的药物组合物及其预防应用。

ISSN1008-4274

1. 一种分离的DNA片段，编码EBV gp350蛋白或截短形式的EBV gp350蛋白，其在跨膜区具有缺失，产生分泌产物、跨膜区缺失和剩余的C末端，和/或有或没有信号片段缺失，所述DNA片段在一个或多个剪接位点具有突变，防止gp220 mRNA转录物的形成。

2. 权利要求1的DNA片段，其中所述在一个或几个剪接位点的突变是供体剪接位点的突变。

3. 权利要求1的DNA片段，其中所述在一个或几个剪接位点的突变是受体剪接位点的突变。

4. 权利要求1的DNA片段，其中所述在一个或几个剪接位点的突变是在供体剪接位点和受体剪接位点的突变。

5. 权利要求4的DNA片段，其中编码SEQ ID NO: 18的501位丝氨酸密码子的至少一个天然核苷酸被非天然核苷酸取代，编码SEQ ID NO: 18的698位甘氨酸密码子的至少一个天然核苷酸被非天然核苷酸取代。

6. 权利要求1的DNA片段，其中DNA编码EBV GP350的截短形式，其在跨膜区具有至少8个氨基酸的缺失，产生分泌产物。

7. 权利要求6的DNA片段，其中所述在一个和几个剪接位点的突变是供体剪接位点的突变。

8. 权利要求6的DNA片段，其中所述在一个和几个剪接位点的突变是受体剪接位点的突变。

9. 权利要求1的DNA片段，其中DNA编码与选自以下的片段相同的氨基酸片段:

(a) SEQ ID NO: 18的密码子19－862;

(b) SEQ ID NO: 18的密码子1－862;

(c) SEQ ID NO: 18的密码子19－862和882－907; 以及

(d) SEQ ID NO: 18的密码子1－862和882－907。

10. 含有权利要求3、5、6、7、8或9的DNA片段的载体。

11. 用权利要求3、5、6、7、8或9的DNA片段转化的宿主细胞，所述DNA片段与能指导所述DNA片段复制和表达的表达控制片段可操作连接。

12. 制备 gp350 蛋白的方法，包括在合适的培养基中培养权利要求 11 的宿主细胞并从所述培养基细胞中分离 gp350 蛋白。

13. 一种均一的 EBV gp350 蛋白，由一种 DNA 片段编码，所述 DNA 片段包括跨膜区缺失和无功能剪接位点，从而 EBV gp350 以可溶的胞外区和胞质区的融合体形式表达，没有 EBV gp220。

14. 一种药物组合物，包括与药物学上可接受的载体混合的权利要求 13 的均一的 EBV gp350 蛋白。

15. 权利要求 13 的 EBV gp350 蛋白，其中跨膜缺失包括 SEQ ID NO: 18 ~~图 1~~ 所示 Ser_{860} 到 Ala_{881} 的缺失（SEQ ID NO: 18）。

16. 含有与药物学上可接受的载体混合的均一 EBV gp350 的组合物，所述 gp350 是从表达和分泌 EBV gp350 的宿主细胞的培养基中分离的，所述宿主细胞向培养基中分泌 EBV gp350 蛋白，所述蛋白由含有缺失跨膜区和无功能剪接位点的 DNA 片段编码，因而表达为可溶形式的胞外区和胞质区的融合体，没有 EBV gp220。

gp350/220的非剪接变异体

疱疹病毒组的一个成员，非洲淋巴细胞瘤病毒(EBV)，引起人的传染性单核细胞增多。该疾病侵袭 90 %以上的人群。健康分析者估测在美国每年该疾病的花费是100百万。该病毒主要是通过与排出该病毒的个体交换唾液而传播的。被 EBV 感染的儿童大多数无症状或者有非常微弱的症状，而青少年和成年人感染后发展成为典型的传染性单核细胞增多，特征为发烧，咽炎和腺病。已经感染过病毒的人群在其余生中保留抗 EBV 抗体，并且因此对进一步的感染有免疫力。目前没有商品 EBV 疫苗出售。

除了感染特性以外，已经证明 EBV 将淋巴细胞转化为快速分裂细胞因此牵涉几种不同的淋巴瘤包括伯基特淋巴瘤和口腔毛细胞粘膜白斑病中。已经在鼻咽肿瘤的组织样本中检测到 EBV 。据估测世界范围内有80,000例鼻咽癌发生并且在中国的少数民族人群中更流行。

对于 EBV 的活的，减毒疫苗的研制已经并且仍然是个问题。由于存在与 EBV 相关的潜在的致癌特性，研究者难于得到使用活疫苗的方法。本发明通过创制制备亚单位疫苗的方法和组合物而克服了与活疫苗研制有关的问题，所述方法和组合物不需要使用潜在的致癌活疫苗。亚单位疫苗使用一种或多种来自于激发免疫应答和授予免疫性的病毒的抗原蛋白。

两种较重要的抗原性 EBV 蛋白是糖蛋白 gp350/300 和 gp220/200，它们构成了病毒包膜部分和通过与细胞膜蛋白 CD21 发生作用允许病毒颗粒结合到并且加入人靶细胞。参见奈麦如，病毒杂志 61:1416(1987)。长期来已经将它们选出作为亚单位疫苗候选物但是难于从天然来源获得有抗原性的活性蛋白以及由于重组生产的产量较低妨碍了研究者和疫苗研制者的工作。在文献中这些蛋白质有不同的分子量范围(其中一种蛋白质的分子量为 350 或 300 千道尔顿(kD)和另一种蛋白质的分子量为 220 或 200kDs)。在本文中将 gp350 或 300 蛋白质称作为 gp350 蛋白质和将 gp220 或 200 蛋白质称作为 gp200 蛋白质。总之，两种蛋白质在本文中称为 gp350/220 蛋白质。

可选择一种经剪接的单基因编码并且导致产生 gp350 和 gp220

mRNA 转录产物；已知没有天然存在的在 gp350/220 基因剪接位点处的变异体该基因产生两种表达产物， gp350 和 gp220 蛋白质。gp350/220DNA 序列的开放读框是 2721 碱基对。该整个阅读框架编码 gp350 的 907 个氨基酸。参见授权给凯夫的 U.S. 专利 No.4,707,358(1987)。剪接形式的阅读框架具有 2130 个碱基并且转译为 gp220 蛋白质，一种 710 个氨基酸的序列。gp350 蛋白质和 gp220 蛋白质的理论分子量是分别为 95kD 和 70kD 。经检测，表达的 gp350 蛋白质和 gp220 蛋白质的分子量是不定的并且分别为约 350 千道尔顿和 220 千道尔顿(kD)。该蛋白质的广泛糖基化是预测的和实际的分子量有差异的原因。对于任何一个细胞，以约 6： 1 到 1： 1 的摩尔比产生 gp350 蛋白质和 gp220 蛋白质。例如，对于由 EBV 持久感染的 B95 － 8 细胞，该比例似乎是不定的并且有时接近于 6： 1 的范围。参见，米勒，美国国家科学院院报， 69:383(1972)。

类似地，至今为止重组生产的这些蛋白质通常得到所产生的 gp350 蛋白质和 gp220 蛋白质的混合物。迄今为止，已经在大鼠垂体，中国仓鼠卵巢 VERO(非洲绿猴肾)细胞，以及酵母细胞中表达了 gp350/220 蛋白质。参见王,J.Virol.61:1796(1982),Motz,Gene 44:353(1986)和艾米尼，病毒学 166:387(1988)。已经利用牛乳头瘤病毒表达系统在鼠成纤维细胞中制备 gp350/220 蛋白质。参见， Madej,Vaccine 10:777(1992)。还已经利用痘苗病毒的实验室和疫苗株表达 gp350/220 蛋白质。经修饰的重组形式的 EBVgp350/220 DNA 和蛋白质是本领域内已知的。特别是已经制备了失去跨膜序列的 gp350/220 基因的截短的重组构建体。所述构建体仍产生两种 gp350 和 gp220 的混合物，但是失去了允许蛋白质分泌的跨膜区域。参见，费纳提，基因病毒学杂志 73:449(1992)和梅者，疫苗 10:777(1992)。也已经在大肠杆菌中制备和表达了含有 gp350/220 的各种重组生产的限制性片段和融合蛋白。参见 1991 年 7 月 24 公开的 EP 专利公开 0173254 。

因此,迄今为止涉及 gp350/220 的对 EBV 的研究已经集中于使天然 gp350/220 序列获得有效表达或者集中于失去跨膜区的经修饰的序列，结果产生两种不同剪接形式的天然或失去跨膜区的蛋白质的混合物，或集中于生产与β半乳糖苷酶融合蛋白形式的抗原决定族片段序列。

部分纯化的 gp350/220 制剂是已知的。参见，

Finerty,J.Gen.Virology 73:449(1992)(重组生产的，部分纯化的)。对于天然 gp350/220 蛋白质，在大多数情况下，纯化步骤导致蛋白质的抗原性失活，使之不适合用于亚单位疫苗中。但是，在科学文献中已经报道了从天然(即非重组)来源获得的高度纯化的具抗原活性的 gp350/220 蛋白质制剂。参见， Dvaid,J.Immunol.Methods 108:231(1988)。将其它的重组疫苗病毒表达的 gp350/220 蛋白质用于接种于 cottontop tamarin ，使之抵抗 EBV 诱导的淋巴细胞瘤。参见， Morgan,J.Med.Virology 25:189(1988),Mackett,EMBO J.4:3229(1985)和 Mackett,VACCINES'86,pp293(Lerner RA,Chanock RM,Brown F Eds.,1986,Cold Spring Harbor Laboratory)。但是，目前还没有对病毒 gp350/220 DNA 序列进行工程化以便能够表达该基因的两种剪接方式的任何一种，从而能够并且确保生产纯化的 gp350 或 gp220 蛋白质。还没有制备出能表达 gp350 或 gp220 蛋白质的任何一种的重组病毒或突变病毒。

通常，剪接位点有利于使前 mRNA 分子加工为 mRNA 。对于多瘤病毒，要求剪接位点有利于有效地积累晚期 mRNA 。多瘤病毒转录物中的 3' 或 5' 剪接位点的变更降低或完全阻碍 mRNA 的积累。参见， Treisman,Nature 292:595(1981)。在 SV40 病毒中，可切除的间插序列有利于从核中输出 mRNA 和 mRNA 在核中的稳定性。并且由于这些内含子/外显子连接序列有利于小的，核的， RNP 颗粒结合，因此认为剪接前的 mRNA 不能正确地与加工途径相连接。已经证明在外显子/内含子剪接位点处的点突变降低了外显子/内含子裂解和可以干扰前－mRNA 加工，核运输和稳定。参见， Ryu,病毒杂志 63:4386(1989)和 Gross,Nature 286:634(1980)。

因此，直到本发明，剪接位点的修饰对有 EBVgp350/220 序列编码的蛋白质的功能表达和抗原性的影响充其量是未知的并且是不可预测的。

其它的背景文献包括如下所述。在 Straus,Annal of Int.Med. 118:45(1993)中总体上评述 EBV 生物学和疾病。在 Baer, Nature 310:207(1984)中描述了 EBV BLLFI 开放读框。由 Beisel,病毒杂志 54:665(1985)和 Biggin,EMBO J.3:1083(1984)d 的论文和在 Kieff 等人(1987)的美国专利 No.4,707,358 中描述了非洲淋巴细胞瘤病毒 gp350/220

DNA 和氨基酸序列。在 Lees,Virology 195:578(1993)中公开了非洲淋巴细胞瘤病毒 A 型和 B 型编码 gp350/220 糖蛋白的 DNA 序列的比较结果。在 Thorley-Lawson,Proc. Natl.Acad.Sci.77:5307(1980)中公开了展示抗 EBV 的 gp350/220 糖蛋白的中和活性的单克隆抗体。最近，在 Mount,Nucleic Acids Res.10:459(1982)中公开了供体和受体剪接位点的剪接位点共有序列。

一方面本发明提供了 EBV gp350/220 DNA 序列的非剪接变异体。本发明的 DNA 序列包括一种编码表达的同源 gp350 蛋白质的分离的 DNA 序列。编码 gp350 蛋白质的 DNA 序列的特征在于含有相同于或基本相同于图 1 的核苷酸序列，其中供体和受体剪接位点处的天然核苷酸被非天然核苷酸或其片段取代。该 DNA 序列包括编码序列两侧的 5’和 3’非编码序列以及进一步包括氨基末端信号序列。附图 1 描述了非编码序列和用星号指出了假设的信号序列。但是，应该明白本发明的 DNA 序列可以排除一些或所有侧翼或信号序列。通过在编码 gp350/220 蛋白质的基因 DNA 序列的供体和受体剪接位点处导入突变可产生本发明的非剪接变异 DNA 序列。这种方法排除了 gp220 蛋白质的生产以便仅生产 gp350 蛋白质。

因此，另一方面本发明包括均一的 gp350 蛋白质，和通过在合适的表达控制序列控制的条件下在合适的原核或真核宿主细胞中表达 EBV gp350/220 DNA 序列的非剪接的变异体而制备蛋白质的方法。对于本文中使用的术语 gp350 蛋白质，均一是没有或基本上没有 gp220 蛋白质。我们注意到在哺乳动物或昆虫细胞中重组产生的均一 gp350 蛋白质在迄今为止的科学文献中没有报道过。

在另一方面，提供了另外具有导致产物分泌的缺失的均一 gp350 蛋白质。所述缺失包括去除跨膜区或去除跨膜区和保留 C－末端的 gp350 。所述附加修饰的 DNA 序列和由其编码的蛋白质是本发明的又一方面。

还提供了一种重组 DNA 分子，包括载体 DNA 和编码均一 gp350 蛋白质的 DNA 序列。该 DNA 分子提供了与合适的调节序列可操作结合的 gp350 序列，所述调节序列能够在选定的宿主细胞中指导均一 gp350 的复制和表达。本发明还提供了用于表达重组均一 gp350 的用所述 DNA 分子转化的宿主细胞。

本发明的另一方面涉及所述的 DNA 分子和转化的宿主细胞，涉及用于生产重组均一 gp350 蛋白质或其片段的新方法。在该方法中在合适的允许所述重组 DNA 表达的条件下培养用一种与能控制该蛋白质表达的合适调节或表达控制序列可操作连接的 DNA 序列(或如上文中描述的重组 DNA 分子)转化的细胞系，所述 DNA 序列编码均一 gp350 蛋白质或其片段。然后采用合适的常规手段从宿主细胞或培养基中收集所表达的蛋白质。该方法可以使用许多已知的细胞作为宿主细胞；目前优选的是哺乳动物细胞和昆虫细胞。

本发明的 DNA 序列和蛋白质可用于生产具有 EBV 抗原决定簇的治疗和免疫原化合物。所述化合物可用于亚单位疫苗，用于预防治疗和预防 EBV 相关的疾病，例如传染性单核细胞，伯基特淋巴瘤或鼻咽癌。因此，另一方面本发明包括用于预防和治疗人的 EBV 相关状况和疾病例如传染性单核细胞，伯基特淋巴瘤或鼻咽癌的治疗和/或免疫原的药物组合物。所述治疗和/或免疫原的药物组合物包括与药物学上可接受的载体例如氢氧化铝，盐水和磷酸缓冲盐水(如本发明已知的)混合的能有效诱导免疫原性量的一种或多种本发明的均一 gp350 蛋白质。“诱导免疫原性的”是指其量足以刺激哺乳动物产生抗 EBV 的抗体。另一种可选择的方法，活性形式可以以含有脂质体的聚合体的形式给药。对于预防应用，所述药物组合物可以配制成为施用于人患者的亚单位疫苗。可以用足于刺激病人形成抗体的剂量免疫接种患者；并且在六个月或一年之后再接种。

因此本发明的又一方面是通过给患者特别是人患者施用与合适载体结合的能有效诱导免疫原性量的均一 gp350 蛋白质而治疗 EBV 相关疾病和状态的方法。 本发明的又一方面是通过给患者施用与合适载体结合的能有效诱导免疫原性量的均一 gp350 蛋白质而刺激抗 EBV 的免疫应答。

附图 1 描述了 gp350/220 的 DNA 和氨基酸序列(Beisel,病毒杂志 54:665(1985))。指明了供体和受体剪接位点。用水平箭头描述跨膜区并且用星号表示假设的信号序列的末端。在左边显示核苷酸的编号；在右边显示氨基酸的编号。

附图 2 描述了 gp350 缺失和定点突变的结构。标明为 pMDTM 和 pMSTOP 的质粒图谱用于举例说明本发明的非剪接的 gp350/220 变异

体。在(A)部分中，近似地按比例显示了 gp350 蛋白质的线性模型以及与编码克隆， BLLF1(见下文). 指明了蛋白质上的 N－末端信号肽序列(SS)和跨膜区(TM)并且在基因图上指明了重要的限制性图谱。将 gp350 基因克隆于两个片段， HindIII/BfaI BLSH1 片段和 BanI/HindIII BLSH2 片段中。利用聚合酶链反应从所指明的 BLLF1 区制备 SCYT 。在(B)中，显示了 pDTM,pSTOP,pMDTM,和 pMSTOP 的克隆方案(质粒未按比例显示)。实施例 1 和 2 详细描述了克隆步骤。用相应的限制性位点，使用的克隆载体和 gp350 基因片段标记质粒图谱。用星号显示 pMDTM,和 pMSTOP 的剪接位点。

附图 3 描述了由 SDS－PAGE 分析的由 pMDTM 克隆产生的均一 gp350 蛋白质的免疫沉淀的结果。分泌截短形式的 gp350/220 蛋白质的阳性对照(GH3Δ19)细胞和阴性对照细胞以及几种 pMDTM 克隆用 ^{35}S－甲硫氨酸在代谢过程进行标记 5.5 小时；从得到的组织培养上清液免疫沉淀出均一的 gp350 蛋白质。对于每一种细胞类型，在 5 % SDS－PAGE(聚丙烯酰胺凝胶电泳)上将标记的组织培养上清液(S)和 gp350 /220 沉淀(Ip)样品进行电泳。在左边指示分子量标记物的位置。

附图 4 描述了来自 CHO 细胞表达的 pMDTM 克隆的蛋白质，如实施例 3.4 的 Northern 印迹分析结果。

公开了包含编码非剪接的 gp350 蛋白质变异体的克隆的 EBV DNA 序列组合物和方法。正如所解释的，本文中所述非剪接变异体是指均一 gp350 蛋白质。通常，当 gp350/220 基因在哺乳动物细胞中表达时由于基因的 RNA 剪接作用产生两种基因产物 gp350 和 gp220 。本发明仅允许产生一种基因产物 gp350 。本发明涉及从 gp350 基因中去除一些或整个 RNA 剪接位点信号并且在合适的宿主细胞中表达该基因。当在哺乳动物细胞中表达 gp350/220 基因时在该基因中导入突变以便阻止产生该蛋白质的 220kD 形式。结果仅产生编码 gp350 的 mRNA 转录物。通过利用 gp350/220 基因非剪接变异体而去除 gp220 的表达导致相对 gp220 而言 gp350 的产生增加。 gp220 的产生不是生产有效抗 EBV 疫苗所必须的，因为 gp350 含有在 gp220 中发现的所有有效的抗原位点。

因此，本发明的一方面提供了编码基本上相同于 gp350 的多肽的 DNA 序列，所不同的是通过用非天然核苷酸取代天然核苷酸已经修饰了编码氨基酸 501 的供体剪接位点密码子和编码氨基酸 698 的受体剪接位

点密码子。优选地是用非天然核苷酸取代天然核苷酸以便保留相同的氨基酸序列。特别是，在该实施例中，在供体剪接位点的天然核苷酸AAGT(核苷酸 1500 到 1504)和 GG 受体剪接位点侧翼的天然核苷酸 A 和 T(核苷酸 2091 和 2094)分别用核苷酸 GTCA 和 T 和 A 取代。因此，在供体位点取代的结果保留第 500 位的氨基酸谷氨酸和第 501 位的丝氨酸是相同的。另外，在受体位点取代的结果保留第 697 位的氨基酸苏氨酸和第 698 位的甘氨酸是相同的。

类似地，采用本领域内已知的在下文中将详细描述的接受可以方便地完成除特别列举以外的取代。

因此，本发明一方面包括均一的 gp350 蛋白质。均一 gp350 蛋白质的进一步的特征在于其氨基酸序列基本上相同于附图 1 显示的氨基酸 1 - 907 位，氨基酸 1 - 862 位或氨基酸 1 - 907 位并且除去氨基酸 863 - 881 位的序列，各带有或没有 N - 末端的 18 个氨基酸的信号序列。另外，提供了均一的 gp350 蛋白质的类似物并且包含突变体，其中氨基酸序列发生变异但是保留了抗原活性并且与均一的 gp350 蛋白质的相应区域相比优选的是具有至少 80 %，更优选的是具有 90 %，最优选的是具有 95 % 的同源性。实施例包括与附图 1 的氨基酸序列相比具有最小变异特别是保守的氨基酸取代的蛋白质和多肽；保守的取代是那些与其侧链相关的氨基酸族之内发生的取代。通常遗传上编码的氨基酸可划分为四个族：(1)酸性 = 天冬氨酸，谷氨酸；(2)碱性 = 赖氨酸，精氨酸，组氨酸；(3)非极性 = 丙氨酸，缬氨酸，亮氨酸，异亮氨酸，脯氨酸，苯丙氨酸，甲硫氨酸，色氨酸；和(4)未带电荷的极性 = 甘氨酸，天冬酰胺，谷氨酰胺，半胱氨酸，丝氨酸，苏氨酸，酪氨酸。有时将苯丙氨酸，色氨酸和酪氨酸一起划分为芳香族氨基酸。例如，有理由预期用结构相关的氨基酸隔离地取代亮氨酸或类似的保守取代一种氨基酸对抗原活性或功能特性没有巨大影响。

本发明具有 gp350 的纯化较简单的优点。由于 gp350 和 gp220 具有类似的生化特性，在 gp350 制备时经常共纯化 gp220 . 仅表达 gp350/220 基因的非剪接变异体的细胞使蛋白质的纯化简单化。这将降低生产 gp350 的费用。本发明也使对用于 gp350 纯化的起始材料进行生化定性变为简单。由于仅存在一个品种，可在不考虑存在第二个品种的情况下完成蛋白质含量分析和氨基酸序列分析。

另外本发明具有增加 gp350 产生的优点。阻止 gp350 基因剪接可将细胞从双重产生 gp350 和 gp220 变为仅产生 gp350。在一些细胞中，估测 gp220 的浓度为 gp350 浓度的 30 % - 100 %。由于去除了基因的剪接，gp350 的产生随着 gp220 产生的损失而提高。

由 Beisel,病毒杂志 54:665(1985)和 Biggin,EMBO J.3:1083(1984)描述了 gp350/220 基因的 DNA 序列并且描述于附图 1 中。该基因是 2721 碱基的开放读框，编码 907 个氨基酸并且产生约 95kD 的一级转译产物。预定的和实际的值之间的差异代表该蛋白质被广泛地糖基化。将 591 个碱基(编码 197 个氨基酸)剪接以便产生 gp220。gp350/220 基因产物的表观分子量随着所使用的检测系统的类型，不同类型的细胞中使用的糖基化位点，转译后的不同加工方法或选择性基因的突变而变化。对于不同的 gp350/220 基因非剪接位点变异体的产物，所测定的值是不同的，但是术语“均一 gp350 蛋白质”包括非剪接位点变异体的基因产物，非强制性地含有其它的缺失或突变例如 C - 末端缺失和/或 本文中公开的跨膜区修饰。术语“gp220 蛋白质”是指具有分子量约 220kD 的另一种剪接的 gp350/220 基因产物。通过将 gp350/220 基因与基于其它基因，主要是真核生物体的基因的共有的供体和受体剪接序列进行比较可识别 gp350/220 基因中的剪接位点。由 Mount,Nucleic Acids Res. 10:459(1982)从研究其它基因的剪接位点得出的共有序列是:

供体: $\frac{A}{C}AG/G^{*}T^{*}\frac{A}{G}AGT$

受体: $\frac{T}{C}N_{11}\frac{C}{T}A^{*}G^{*}/G$

上述带星号的碱基代表在所有的剪接位点 100 % 出现的碱基(高度保守)。有两个碱基或一个碱基的位置代表保守位置(非强调的位置)。斜线位置表示实际的剪接位点。

在 gp350/220 基因中通过对 gp350/220 基因的 DNA 序列测序(Biggin,EMBO J.3:1083(1984))显示供体剪接位点出现于核苷酸 1501 之后，受体剪接出现于核苷酸 2092 之后。(本文中使用的和附图 1 使用的编号对应于 Biggin 的编号)。在 EBV 的 B 型菌株该剪接位点出现于相应

的基因区(供体剪接位点出现于 A_{1501} 之后，受体剪接出现于 A_{2029} 之后)。本发明包括利用 A 或 B 菌株或另一种 EBV 菌株的剪接位点以便从 gp350/220 基因产生单一的 mRNA 而制备的组合物。在该实施例中使用了 B95-8 菌株的 A 型病毒形式的 DNA 序列，尽管使用 B 型菌株的 DNA 序列是相同的，因为 A 和 B 型菌株的转译基因产物 98 % 是相同的。B 菌株失去了 507 - 520 位的氨基酸和 570 - 576 位的氨基酸。由于 A 型菌株含有所有可能的 gp350 抗原位点因而使用 A 型菌株。另一种方法，根据本文的观点使用具有菌株特异性序列的 EBV gp350/220 以便产生 EBV 菌株特异性的均一 gp350 蛋白质，它具有特异于特定菌株的免疫特性并且因此用于免疫和/ 或治疗组合物中以便预防或治疗菌株特异性 EBV 相关疾病。表 1 显示供体和受体剪接位点的野生型核苷酸和氨基酸序列。

为了预防 gp350/220 基因的 RNA 剪接，将突变导入到 gp350/220 基因的核酸序列以便取代 RNA 剪接位点的有关碱基对。为了使剪接位点失去功能，优选地，在形成供体或受体位点框架的两个高度保守碱基中的至少一个碱基应该被非保守碱基取代，更优选地至少两个高度保守碱基应该被突变为非保守碱基。其它保守碱基，剪接位点之外的两个以上碱基也可以被非保守剪接位点碱基取代以便进一步降低剪接位点的识别。将供体和受体位点改变以便破坏剪接机理。优选地，供体和受体在四个高度保守剪接位点碱基位置之一至少各含有一种变化，并且优选地，至少在四个高度保守剪接位点碱基位置之二各含有两种变化。如果期望一个剪接位点没有突变以便保持野生型氨基酸序列，那么优选地将至少两个突变导入到其它剪接位点。

在 gp350/220 剪接位点的突变可以使随后表达的氨基酸序列发生变化。优选地所述变化应该是保守的氨基酸取代。与氨基酸序列中非保守的取代相反，保守的氨基酸取代将帮助保存抗原位点。可以制备保守的氨基酸变化，只要该碱基的变化(几个碱基变化)导致非变异的供体/受体碱基发生合适的变化。例如，利用除 GGU 以外的特异于 Gly 的密码子可以在供体剪接位点以 Gly 取代 Ser_{501}(利用 GGU 会保存 G 核苷酸并且不会在剪接信号产生所需的 GT 取代)。另外，在受体剪接位点，以 Gly_{698} 取代 Ala 将是保守的变化，但是由于所有的 Ala 密码子是以高度保守的 G 核苷酸起始的，这不会产生所需的取代。虽然脯氨酸也可能是一种保

守的氨基酸变化，但是不能用脯氨酸取代野生型氨基酸，因为它会导致蛋白质的三级结构发生修饰，从而掩饰一个或多个 gp350 抗原位点。表 1 显示野生型序列中的可接受的保守的氨基酸取代。在表 1 的底部是用保守的氨基酸序列突变的实施例。

表 1

	Donor			Acceptor	
			Wild-type Sequences		
	splice			splice	
	GAA	A|GT		ACA	G|GT
	Glu	Ser_{501}		Thr	Gly_{698}
	↓	↓	Conservative a.a. changes	↓	↓
	Asn	Ala		Ala	Ser
	Asp	Gly		Gly	Thr
	Gln	Thr		Ser	
ex.:	GAC	ACA		TCG	TCT
	Asp	Thr_{501}		Ser	Ser_{698}

虽然本发明的一方面包含 gp350/220 的非剪接变异体，但是其它的 gp350/220 的编码序列的突变体也是令人满意的。为了生产可溶性 gp350 蛋白质(“可溶性蛋白质”是在溶液中游离的或与膜结合但是不与膜整合)，例如，以便避免由于表达整合膜蛋白形式的全长 gp350 而产生的细胞毒性问题，通过缺失整个或部分其 DNA 编码序列而对 gp350 的膜跨越区(已知也称为跨膜区)进行修饰。 gp350/220 的膜跨越区包含 861 位(甲硫氨酸) - 881 位(丙氨酸)氨基酸。参见， Beisel, 病毒杂志 54:665(1985)。优选地，缺失跨膜区的至少 8 个氨基酸，更优选地缺失至少 12 个氨基酸，最优选地缺失 18 和 21 之间的氨基酸。因此，另一方面，本发明提供了 gp350/220 DNA 的非剪接变异体和/或 gp350 均一蛋白质，特别是在 gp350/220 DNA 和/或 gp350 均一蛋白质的跨膜区至少有一个缺失，从而导致可溶性均一 gp350 蛋白质的表达。

除了缺失整个或部分非剪接 gp350/220 变异体的跨膜区，如本文描述的根据本发明也可以缺失跨膜区后面的含有 881 - 907 位的氨基酸的 C - 末端序列。因此另一方面本发明包含进一步由缺失整个或部分 DNA 编码序列和/或氨基酸序列(gp350/220 的跨膜区)，并且甚至进一步由缺

失其余的 C - 末端 DNA 和/或 gp350/220 的氨基酸序列而修饰的 gp350/220 DNA 的非剪接变异体和/或 gp350 均一蛋白质。

因此，另一方面本发明包含编码本发明的均一 gp350 蛋白质的非剪接的 DNA 序列变异体。所述 DNA 序列包含编码 1 - 907 位氨基酸的附图 1 的 DNA 序列并且进一步含有本文所述的核苷酸取代以便去除供体和受体剪接位点。所述 DNA 序列非强制性地含有截短的 DNA 序列，其中缺失了编码整个或部分跨膜区和包含 861 - 881 位的 C 末端的核苷酸，并且含有缺失变异体，其中缺失了编码含有 861 - 881 位的整个或部分跨膜区的核苷酸。编码均一 gp350 蛋白质的本发明的 DNA 序列也含有能在合适的严谨条件下，或要不是遗传密码子简并能在所述条件下与附图 1 的分离出的 DNA 序列杂交的 DNA。因此，基于等位变异，种变异或故意修饰，本发明的 DNA 序列在非编码序列，信号序列或编码序列中可以含有修饰。

利用本领域内已知的方法可以构建本文中公开的这些非剪接变异 gp350/220 DNA 序列。可以采用重组方法，或者利用已知方法表达本发明的修饰 DNA 序列以便生产本发明的均一 gp350 蛋白质。可将所述重组蛋白质纯化并且掺入到药物组合物中用于预防治疗和防止 EBV 相关疾病。

可利用本领域内已知的合适表达控制系统在不同的细胞类型中重组表达本发明的非剪接的变异 gp350/220 DNA。本领域内已知和可行的合适的细胞包括但不限于酵母细胞例如啤酒酵母，细菌细胞例如大肠杆菌和枯草芽孢杆菌和哺乳动物细胞例如 GH3，CHO，NSO，MDCK 和 C - 127 细胞。根据与所使用的细胞类型和表达控制系统的相容性选择适应于细胞类型的载体。优选地，使用表达 gp350/220 基因的分泌产物的细胞和载体。通常例如，利用已经用常规技术修饰而含有表达所需蛋白质的 DNA 序列的 pBR322 的衍生物转化大肠杆菌，在本发明中其中所述的 DNA 序列是有或没有编码 C - 末端和/或膜跨越区的序列的 EBV gp350 的非剪接变异体。pBR322 含有氨苄青霉素和四环素抗性的基因，用作为标记物。参见,Bolivar,Gene 2:95(1977)。通常使用的表达控制序列即用于转录起始的启动子和非强制性包括的操纵基因或加强子，包括β - 内酰胺酶和 lac 启动子系统(参见 Chang,Nature 198:1056(1977)),色氨酸启动子系统(参见 Goeddel,Nucleic Acids

Res.8:4057(1980))和λ衍生的 PL 启动子和 N －基因核糖体结合位点(参见 Shimatake,Nature 292:128(1981))。但是，可以使用与原核宿主细胞相容的可行的启动子系统或表达控制系统。在授权给 Itakura(1982)的美国专利 No.4,356,270，授权给 Riggs(1984)的 4,431,739 和授权给 Rutter(1984)的 4,440,859 中列举了其它的宿主细胞，质粒和表达载体。

也可以将昆虫细胞用作为宿主细胞以便利用昆虫细胞进行表达。对于在昆虫细胞表达的情况，通常表达系统的成分包括经常为细菌质粒的转移载体和一种具有与转移载体中的杆状病毒特异性片段同源的野生型杆状病毒(从而允许异源基因同源重组到杆状病毒基因组中)以及合适的昆虫宿主细胞和生长培养基，所述转移载体含有杆状病毒基因组的一个片段和用于插入异源基因的或待表达便利的限制性位点。

目前，将外源基因导入到 AcNPV 中的最常用的转移载体是 pAc373。也已经设计了许多本领域内技术人员已知的其它载体。例如包括 pVL985(它将多角体蛋白起始密码子 ATG 改变为 ATT，并且在 ATT 的下游的 32 碱基对处导入了 BamHI 克隆位点；参见 Luckow 和 Summers,Viorlogy(1989)17:31)。

通常该质粒还含有多角体蛋白聚腺苷酸化信号(Miller 等人(1980)Ann.Rev.Microbiol.,42:177)和原核的氨苄青霉素抗性(amp)基因和用于在大肠杆菌中选择和繁殖的复制原点。

杆状病毒转移载体通常含有杆状病毒启动子。杆状病毒启动子是能够结合杆状病毒 RNA 聚合酶并且启动编码序列(例如结构基因)下游(5'-3')转录成为 mRNA 的任何 DNA 序列。启动子具有转录起始区，通常置于接近编码序列 5'末端。通常该转录起始区包括 RNA 聚合酶结合位点和转录起始位点。杆状病毒转移载体也可以具有称为加强子第二结构域，如果存在该加强子，通常远离结构基因。表达可以是被调节的或组成型的。对于昆虫细胞表达技术，参见 EP 专利公开 155 476。

酵母，例如酿酒酵母亦可用作宿主细胞。能得到并使用多种细胞株。另外，适应于酵母表达的质粒载体是已知的，启动子和表达控制系统也是已知的。参见例如，Myanohara，Proc.natl.Acad.Sci.80：1（1983）（PHO5 启动子），EP 专利公开 012873（引导序列），Kurtz，Mol.Cell.Biol.6：142（1986），Ito，J.Bacteriol.153：163（1983）和 Hinnen，Proc.natl.Acad.Sci.75：1929（1979）（转化程序和合适

载体）。

当然也可以将来自多细胞生物体的真核细胞用作为表达编码所需蛋白质和多肽的基因的宿主细胞。有用的宿主细胞系包括 VERO 和 HeLa 细胞，和中国仓鼠卵巢细胞(CHO)。与所述细胞相容的表达载体也是可用的并且典型包括启动子和表达控制序列，例如 SV4 的早期和晚期启动子(参见， Fiers,Nature273:113(1978))和来自多瘤病毒，腺病毒 2，牛乳头状瘤病毒或鸟肉瘤病毒的启动子。在授权给 Mather(1992)的美国专利申请 No.5,122,469，授权给 Axel(1983)的 4,399,216,授权给 Axel(1987)的 4,634,665，授权给 Levinson(1987)的 4,713,339，授权给 Ringold(1987)的 4,656,134，授权给 Kellems(1989)的 4,822,736，授权给 Fiers(1989)的 4,874,702 中列举了宿主细胞，启动子，选择性标记和技术。

利用适合于所述细胞的标准技术，例如在 Cohen Proc.Natl.Acad.Sci.69:2110(1972)中公开的用于原核细胞的 $CaCl_2$ 处理法和在 Graham,Virology 52:546(1978)公开的用于哺乳动物细胞的 $CaPO_4$ 沉淀法可完成合适的宿主细胞的转化。按照在 Hsiao,Proc.Natl.Acad.Sci.76:3829(1979)或如在 Klebe,Gene25:333(1983)公开的方法完成酵母的转化。

利用本领域内已知的常规连接法和限制性技术可构建含有非剪接变异 gp350 序列(有或没有在本文中公开的导致 C－末端和/或膜跨越区的其它修饰)的合适的载体。通过在标准条件下，特别是通常由限制性酶制造商提供的条件下用合适的限制性酶处理而进行位点特异性 DNA 裂解。利用标准技术进行聚丙烯酰胺凝胶或琼脂糖凝胶电泳以便按大小将切割的片段分离，并且通过在四个脱氧核苷酸三磷酸存在下用大肠杆菌聚合酶 I 的 Klenow 片段处理使片段的末端成平齐末端。用 S1 核酸酶处理将单链部分水解。利用例如本领域内已知的氨基亚磷酸二乙酯（ diethyl phospho anidite ）方法合成寡核苷酸。参见，美国专利 No.4,415,732(1983)。在标准条件和温度下利用 T4 DNA 连接酶完成连接，通过用连接混合物转化大肠杆菌或 COS 细胞证实正确的连接。利用氨苄青霉素，四环素或其它抗菌素或利用本领域内已知的其它标记选择成功的转化体。

在文献中对所述重组 DNA 技术进行详细解释。参见，例如，

萨姆布鲁克等，分子克隆：实验室手册，第二版（ 1989 ）； DNA 克隆，第 I 及 II 卷（ DN，格拉夫编 1985 ）；寡核苷酸合成（ MJ.盖特编 1984 ）；核酸杂交（ BD 海姆斯编 1984 ）；转录和转译（ BD 海姆斯编 1984 ）；动物细胞培养（ RI 弗莱施尼编 1986 ）； B.坡堡，分子克隆实施指南（ 1984 ）；哺乳动物基因转运载体（ JH 米勒编 1987，冷泉 实验室）；斯科普，蛋白纯化：原则和实践，第二版（ 1987 斯普林格－沃拉登 NY)和免疫学实验手册 Vols I-IV （ DM Weired 1986 ）。

本文中提及的所有出版物引入本文作参考。

因此另一方面本发明包含含有非剪接变异 gp350/220 DNA 序列的载体和宿主细胞和进一步包含通过在能使均一 gp350 蛋白质表达的培养条件下培养所述宿主细胞制造非剪接变异 gp350/220 蛋白质的方法，所述宿主细胞含有携带与表达控制序列连接的非剪接变异 gp350/220 DNA 序列的载体。

利用常规的糖蛋白纯化技术例如但不限于超滤，自由流动的电泳，凝胶过滤层析，亲和层析， SDS-PAGE，差示 NH_4SO_4 沉降，凝集素柱，离子交换柱和本领域内已知的疏水柱，从细胞和培养基成分中纯化表达的均一 gp350 。利用 SDS － PAGE 或凝集素亲和柱最容易纯化 gp350 的小批量的分析制剂，利用液体层析最容易纯化用于疫苗或免疫应答试验的所述小批量制剂。对于用于疫苗组合物的商品显著量的 gp350 的大批量生产，允许的是将超滤，凝胶过滤，离子交换，和疏水作用层析结合应用。

可将本发明的纯化，均一 gp350 蛋白质用于治疗和/或免疫组合物中，用于治疗和预防 EBV 相关的状况和疾病例如，传染性单核细胞增多,伯基特淋巴瘤,鼻咽癌。所述药物组合物包含与药物学上可接受的载体例如助剂/抗原呈递系统例如铝混合的能有效诱导免疫原性量的一种或多种本发明的均一 gp350 蛋白质。也可以使用其它的助剂/抗原呈递系统例如 MF59(Chiron Corp.),QS-21(Cambridge Biotech Corp.)， 3-DMPL(3-脱酰基－单磷酰脂 A)(RibiImmunoChem Research,Inc.)，临床级的不完全弗氏助剂(IFA)， fusogenic 脂质体，水溶性聚合体或 Iscoms(免疫刺激复合物)。其它的典型的药物学上可接受的载体或溶液是氢氧化铝，盐水和磷酸缓冲盐水。该组合物可以是全身给药，优选地以可接受的皮下或肌内溶液形式皮下或肌内给药。通过表皮划破或体腔

接种也可完成接种。所述溶液的制备，包括 pH，等渗性，稳定性等等是本领域内已知的技术。所使用的剂量由医生考虑已知可改变药物作用的各种因子决定的，例如身体状况，体重，性别，食物，疾病的严重性，给药时间和其它临床因子。典型的剂量范围包括约 1μg－约 1000μg 之间的蛋白质。

在实施本发明的治疗方法时，给需要治疗或预防的人患者施用诱导免疫原性有效量的均一 gp350 蛋白质。所施用的每剂量含有约 1μg－约 1000μg 范围内的诱导免疫原性有效量的本发明的组合物。所施用的剂量的数目随着上面介绍的因子而变化。在下面的实施例中进一步描述本发明，这些实施例是为了描述本发明而不是为了限制本发明的范围。

实施例 1

gp350/220 跨膜区和跨越 C 末端的跨膜区的缺失以便制备 pDTM 和 pSTOP

来自 EBV B95－8 菌株(Miller,等人，1972)的 gp350/220 基因以称为 BLLF1 的开放读框形式存在于 BamHI 文库中(Baer,Nature 310:207,1984)。为了制备所需的构建体(显示于附图 2B 中)，将 gp350/220 分两个部分克隆：1)BLSH1，2.3kb 的 HindIII/BfaI3'片段和 2)BLSH2，337b.p.BanI/HindIII 5'片段(附图 2A)。将这些片段克隆到中间载体以便实现 C－末端胞质和/或转膜编码区的缺失。由于在编码 gp350 转膜(TM)区的区域的 5'末端存在 BfaI 位点，可将其用于构建 TM 区的缺失和带有邻近的 C 末端缺失的 TM 区缺失。利用 BfaI，可以制备仅保留 TM 区的两个氨基酸的缺失(表 2)。

1.从 pSTG1 和 pSTG3 构建 pDTM

质粒 pDTM 由失去了整个 TM 编码区的 gp350/220 核酸序列组成。利用两个中间载体 pSTG1 和 pSTG3 可制备该构建体。利用 BLLF1 克隆靶序列可制备 450bp PCR 产物，SYCT，它在 TM 区的 3'末端导入了 BfaI 位点(附图 2)。所使用的 PCR 引物如下：

BfaI

```
Primer 1:   GG · ATC CTA GAC TGC GCC TTT AGG CGT A
BLLF1:          ...  GAC TGC GCC TTT AGG CGT A..
A..A.:          ...  Asp Cys Ala Phe Arg Arg ...
                     ↑___End of TM Region
Primer 2:   GGA TCC TCT GTT CCT TCT GCT CCA GTG
BLLF1:      ... ...     TCT GTT CCT TCT GCT CCA GTG
```

引物 1 的 BfaI 位点用于将 SCYT 的 BfaI/XmaI 片段克隆到 pSTG1。引物 1 的其余部分对应于由克隆 BLLF1 编码的氨基酸序列。引物 2 对应于 gp350/220 开放读框之外的该基因 3’ 端的区域。用 BfaI 和 XmaI 切割 SCYT PCR 片段以便产生 136 碱基对片段，该片段与第二个片段，BLSH1 HindIII/BfaI 片段一起克隆到 pTM11 以制备 pSTG1。跨越 BfaI 位点的测序分析结果显示整个 TM 氨基酸编码区被缺失，仅留下氨基酸 Met 和 Leu(参见表 2)。将第三个 BLLF1 片段，BLSH2 克隆到 pTM11 以制备 pSTG3。将 gp350/220 基因编码序列之外的 16 碱基对 BanI/XbaI

寡核苷酸接头用于将 BLSH2 BanI/HindIII 片段克隆到 pSTG3。将 2.4HindIII/XmaI pSTG1 片段与 0.3 XbaI/HindIII pSTG3 片段一起克隆到 pEE14 载体(Celltech,England)得到整个 pDTM 构建体。

2.利用载体 pSTG2 和 pSTG3 构建 pSTOP

质粒 pSTOP 含有失去了 TM 区和带有邻近的 C 末端胞质区的 TM 区的 gp350/220 基因。为了制备该构建体，制备一个相当于 BfaI 粘性末端 16 碱基对的 BfaI/EcoRI 寡核苷酸接头，它在三个框架中带有终止密码子(底下划线)如下所示：

TAT AGA CTA GTC TAG G

A TCT GAT CAG ATC CTT AA

上排序列的 5’ 延伸臂(TA)是一个 BfaI 限制性位点的粘性末端并且下排序列的 5’ 延伸臂(TTAA)是 EcoRI 粘性末端。将该 16 碱基对接头用于将 BLSH1 HindIII/BfaI 片段克隆到 pTM11 以制备 pSTG2。将 2.3kb pSTG2 HindIII/EcoRI 片段和 pSTG3 0.3kb XbaI/HindIII 片段克隆到 pEE14 以制备 pSTOP。

3.野生型， pDTM 和 pSTOP 序列的 TM 区域的比较

野生型， pDTM 和 pSTOP 3’末端的 gp350 DNA 的寡核苷酸序列和转译的氨基酸序列以及氨基酸序列显示于下文表 2 中。箭头表示野生型跨膜区(TM)的起始点和末端。仅来自跨膜区的两个氨基酸保留于 pDTM 和 pSTOP 中，即 Met_{861} 和 Leu_{862}(也参见附图 1)。在 pSTOP 中终止密码子紧接 Leu_{862} 之后，在 pDTM 中缺失的跨膜区的前面位置标记为“ΔTM”。(在该表中，显示了天然氨基酸)。

表 2

3’ End of gp350 Wild-Type Sequence

...AAC CTC TCC ATG CTA GTA CTG. . . . .GTC ATG GCG GAC TGC GCC

...Asn Leu Ser Met Leu_{862} Val Leu. . . . . Val Met Ala Asp_{882} Cys Ala

↑ ↑

TM start TM end

3’ End of pSTOP

...AAC CTC TCC ATG CTA TAG ACT AGT TCT AGG ...

...Asn Leu Ser Met Leu_{862} STOP

3’ End of pDTM

...AAC CTC TCC ATG CTA GAC TGC GCC...

...Asn Leu Ser Met Leu_{862} Asp_{882} Cys Ala...

∧

ΔTM

实施例 2

去除 gp350/220 基因供体和受体剪接位点以便制备 pMDTM 和 pMSTOP

为了获得均一产生的 gp350 蛋白质，将 gp350/220 剪接位点的高度保守和保守碱基改变。在供体剪接位点改变四个碱基，包括在 100 % 的所有剪接位点中出现的高度保守 GT 对。两个保守的供体位点碱基，AA，被 GT 取代。将高度保守的(不变)供体剪接位点碱基从 GT 改变为 CA 。在受体剪接位点，仅将高度保守的受体剪接位点碱基之一改变以便保存氨基酸序列。如表 2 显示的将第二个保守的受体剪接位点碱基改

变。表 3 概括了在 gp350/220 基因的供体和受体剪接位点改变的碱基。

表 3EBV gp350/220 基因剪接位点的变化

供体剪接

	donor		donor	
Wild-type:	GAA	A↓GT	mutant:	GAG* T*C*A*
	Glu	Ser_{501}		Glu Ser_{501}

Acceptor Splice site:

	acceptor		acceptor	
Wild-type:	ACA	G↓GT	mutant:	ACT* GGA*
	Thr	Gly_{698}		Thr Gly_{698}

在突变序列中用星号标记由寡核苷酸碱基诱变改变的碱基。有箭头表示实际的剪接位点，并且显示了所编码的氨基酸。注意没有由于核苷酸取代改变该氨基酸序列。

利用寡核苷酸介导的诱变可完成对野生型 gp350/220 供体剪接位点和受体剪接位点 DNA 序列的这些核苷酸取代。按照 Zoller,M.E.和 Smith,M.(1983) Methods of Enzymol. 100:468 的方法利用一种修饰的噬菌体载体，M13TAC 产生突变。利用多接头上的 Asp718 和 BamHI 限制性位点，与含有 Asp718 和粘性末端的 19bp 寡核苷酸接头一起将 gp350/220 核苷酸序列的 BamHI/XhoI 片段克隆到质粒 M13TAC 的多接头。利用实施例 1 的质粒 M13DTM 和 M13STOP(附图 2)进行诱变。

将两个 42 - 聚体寡核苷酸，Pr 供体 1 和 Pr 受体 1 用于诱变。将每个设计成为与供体或受体剪接位点中间的 gp350/220 基因序列互补。不与 gp350/220 基因互补的唯一寡核苷酸区域是代表所需突变的碱基。诱变的寡核苷酸如下：

PrDonor1

```
Primer:  GGT CAT GTC GGG GGC CTT TG |A  CTC TGT GCC GTT GTC CCA TGG
                                    ** |*  *
EBV:     GGT CAT GTC GGG GGC CTT AC |T  TTC TGT GCC GTT GTC CCA TGG
```

PrAcceptor1

```
Primer:  CTG TGT TAT ATT TTC ACC TC |C  AGT TGG GTG AGC GGA GGT TAG
                                    *  |  *
EBV:     CTG TGT TAT ATT TTC ACC AC |C  TGT TGG GTG AGC GGA GGT TAG
```

将诱变寡核苷酸序列标记为“引物”，而跨越 gp350/220 基因剪接位点的 DNA 序列标记为“EBV”。用星号表示由诱变改变的碱基。长划线表示剪接位置。

将寡核苷酸 Pr 供体 1 和 Pr 受体 1 与单链克隆的 M13DTM 和 M13STOP 杂交。将 T4 DNA 聚合酶全酶用于生产双链 M13 DNA 并且用双链 DNA 转化大肠杆菌。利用载体 M13TAC，在 X-gal 和异硫丙基半乳糖存在下根据颜色从白色变化为蓝色可鉴别含有所需突变的任何克隆。挑取蓝色噬菌体并且使之生长，将跨越剪接接头的 DNA 测序结果用于最终鉴别突变克隆，称为 M13-MDTM 和 M13 - MSTOP.

在鉴别出含有所需突变之后，从 M13-MDTM 和 M13 - MSTOP 中切割出 BamHI/XhoI 片段并且连接到 pDTM 或 pSTOP 主链以便分别制备 pMDTM 和 pMSTOP 构建体。按照实施例 3 描述将这些构建体转染到 CHO 细胞以便表达非剪接的变异 gp350/220 DNA 序列。

实施例 3

gp350 在 CHO 细胞中的表达

1.gp350/220 基因构建体的转染

用于从哺乳动物细胞生产高含量本发明的均一 gp350 蛋白质的方法涉及构建含有多拷贝异源 gp350 DNA 序列的细胞。将异源 DNA 序列可操作连接到可控制的标志上，在本实施例中该标志是谷氨酰胺合成酶基因，细胞利用甲硫氨酸 sulphoximine 可扩增该基因。

按照下文讨论，根据 Crockett,Bio/Technology 8:662(1990)和在 Celltech Instrution Munual 描述的用于谷氨酰胺合成酶基因扩增的系统可将实施例 2 中制备的 pMDTM 和 pMSTOP 载体转染到 CHO 细胞中。

将 CHO - K1 细胞(ATCC CCR61)保持于补充了 10 % 胎牛血清，100 单位/ml 青霉素，100mg/ml 链霉素，MEM(改良的伊格尔氏培养基)非基本氨基酸，和 1mM 丙酮酸钠(都来自于 JRH Biosciences)的无谷氨酰胺的 EMEM(伊格尔氏基本必要培养基)。该培养基也补充了 60mg/ml 谷氨酸，60mg/ml 天冬酰胺，7mg/ml 腺嘌呤，7mg/ml 鸟嘌呤，7mg/ml 胞嘧啶，7mg/ml 尿嘧啶，和 2.4mg/ml 胸腺嘧啶(都从 Sigma 获得)。在整个转染期间使用该培养基配方，以及按照该配方制备的衍生物。

转染前一天，将 3×10^6 CHO - K1 细胞播种到 10 - cm 培养皿中。在转染当天，每个培养皿用 10ml 无血清培养基洗涤细胞。利用常规的 $CaPO_4$ 沉淀技术得到的质粒 DNA。在 37 ℃ 将各种质粒 DNA 沉淀个 10μg 与 CHO - K1 细胞和 2ml 无血清培养基一起培养 4.5 小时。四个质粒

DNA 转染子个制备三个重复。然后用溶于 HEPES －缓冲盐水中的 15 % 甘油将细胞休克 1.5 分钟。在用无血清培养基洗涤之后，用含有血清的培养基喂养细胞并且培养 24 小时。

第二天将培养基改变使之含有 10 % 渗析过的胎牛血清(JRH Bioscience)并且通过加入 25μM 甲硫氨酸 sulphoximine（ Sigma)进行扩增。每隔 3 － 5 天再用含有甲硫氨酸 sulphoximine 的培养基喂养细胞，直到大约 13 － 14 天之后扩增的克隆大到足于挑取。通过用一个 200μl 无菌的移液管嘴从培养皿中刮下菌落而挑取克隆并且转移到 96 孔平板的一个孔中的没有甲硫氨酸 sulphoximine 的培养基中。1 － 2 天之后用加了 25μM 甲硫氨酸 sulphoximine 的培养基替代孔中 的培养基。4 天之后收集培养上清液利用 ELISA 检测法分析蛋白质产物，如下文中讨论的。

还用 pEE14 对照载体（不含有 EBV 序列）单独转染 CHO 细胞，也挑取 24 个 CHO － pEE14 并且转移到平板上用作为对照。（基于在甲硫氨酸 sulphoximine 上能生存而鉴定对照克隆）。

2.ELISA 检测

转染之后，挑取 241 个 CHO － pMDTM 克隆和 158 个 CHO － pMSTOP 克隆并且使之生长。检测这些克隆的上清液中产生的 gp350 蛋白质。用在 50mM 硼酸钠缓冲液， pH9 中稀释 2000 倍的亲和纯化的兔抗 gp350/220 抗体（抗体 MDPE； Andrew Morgan 赠送） 包被 96 孔平板。将该平板在 37 ℃培养 3 － 4 小时并且利用一个 Nunc Immuno 洗涤器用 PBS + 0.05%吐温 20 洗涤 3 次。在吸干之后，在 37 ℃用溶于 PBS + 0.01% Thimerosal 中的 2 % BSA 培养封闭该平板半小时并且再次洗涤。将转染细胞和对照细胞的上清液加到这些孔中并在 37 ℃培养 2 小时。然后在 37 ℃用在 PBS 洗涤缓冲液中稀释的 1mg/ml 的一级检测抗体，一种抗 gp350/220 的鼠单克隆抗体(抗体 # C65221M： Biodesign International)培养该平板 1 小时。洗涤之后在 37 ℃用溶于 PBS + 0.05%BSA 和 0.01% Thimerosal 中的 0.7μg/ml 抗鼠免疫球蛋白的辣根过氧化物酶结合的羊 F （ ab)$_2$ 片段（人 Ig 吸附的； Biosource International)二级抗体培养该平板 1 小时。在室温下用溶于稳定的过氧化物底物缓冲液（ Pierce Chemicals ）的 ABTS （ Pierce Chemicals ）洗涤该平板并且显色半小时。用 1 % SDS 终止反应，利用 Molecular

Devices Vmax ELISA 平板计数器在405和650nm波长处对该平板进行计数。24个pMDTM和18个pMSTOP克隆检测为分泌gp350的阳性。将显示最强 ELISA 信号的克隆转移到 24 孔平板上并且进一步在 Western Blot 和放射免疫沉淀分析中检测。

3.Western Blot 和放射免疫沉淀分析中检测

在开始筛选时，在 Western Blot 中分析 pMDTM 转染的组织培养上清液的活性。在 5 % SDS - PAGE 凝胶上纯化 CHO 细胞上清液，转移到硝酸纤维素膜过夜，并且用抗 gp350 抗体探测。发现 7 个 pMDTM 克隆在 Western Blot 分析中显示 gp350 阳性。

进一步采用存在 gp220 的放射免疫沉淀方法检测在 Western Blot 分析中显示阳性的 pMDTM 克隆。在 6 个平板孔中将选出的转化 pMDTM 细胞， pEE14 对照和 GH3Δ19 对照细胞（下文描述）培养过夜以便在试验当天有 3/4 的细胞汇合。每个孔大约含有 5×10^6 细胞。为了进行标记，从每个孔中除去培养基并且用 0.7ml 添加了 100μCiS^{35} 甲硫氨酸的无甲硫氨酸 MEM （ 10 % 胎牛血清)替代。在 37 ℃将细胞培养 5.5 小时，然后以 4000rpm 微离心 5 分钟。通过加入 10μlSepharose-蛋白质 A(Sigma)和 20μl 单克隆抗 gp350/220 （抗体 # C65221M ： Biodesign International)在 4 ℃旋转过夜而将上清液中的均一 gp350 蛋白质免疫沉淀。然后在室温下微离心机中以 2000rpm 沉淀该混合物 2 分钟，用数倍体积的磷酸缓冲盐水洗涤 4 次。在最后洗涤之后从沉淀中除去所有的液体并且用 50μl 蛋白质凝胶样品缓冲液替代。将含有沉淀的免疫复合物的样品煮沸 5 分钟并且在 5 % SDSPAGE 上电泳。免疫沉淀物相当于与蛋白质样品缓冲液以 1： 1 混合的组织培养上清液的凝胶样品。将该凝胶干燥并且用 Hyperfilm β-Max(Amersham)放射自显影。

附图 3 显示放射免疫沉淀的 SDSPAGE 分析的放射自显影结果。用作为阳性对照的细胞系是 GH3Δ19 （ Elliot Keiff 赠送； Whang,等人，1987 ）。 GH3Δ19 细胞分泌失去了转膜和 C 末端胞质区的截短形式的 gp350/220 蛋白质。为了用作为阴性对照，单独用 pEE14 载体转染 CHO 细胞并且同时用 pMDTM 转染通过甲硫氨酸 sulphoximine 选择。在附图 3，奇数泳道显示上清液（ “ S ”），偶数泳道显示免疫沉淀（ “ Ip”）。在对照泳道 2，来自 GH3Δ19 对照细胞的沉淀产生约 220 和 350KD 的两个强蛋白质谱带，表明以 1： 1 的比例产生截短的剪接变体 gp350 和

gp220 蛋白质。正如所期望的在泳道 1 这些免疫沉淀谱带集中了放射标记的组织培养上清液（非免疫沉淀样品）。也正如所期望的由于 pEE14 载体不含有 gp350/220 构建体，在阴性对照没有显示任何谱带（泳道 4)。

在泳道 6， 8 和 10 来自 pMDTM 克隆上清液的免疫沉淀的 SDSPAGE 分析结果显示有一个约 350KD 的单一强谱带，相当于对照泳道 2 中的 GH3Δ19 对照的较高分子量种类。但是与 GH3Δ19 对照泳道相反，泳道 6， 8 和 10 没有另一个约 220KD 的强谱带，虽然泳道 8 显示迁移在稍微较低分子量的一个较微弱的谱带。这可能代表一种降解产物，一种共沉淀的细胞产物或者由于错翻译或突变产生的少量 gp220 蛋白，所述突变将缺失的供体和受体剪接位点返回到天然的核苷酸或氨基酸序列。在其它 5 个所检测的 MTDM 重复中发现约 350KD 处的强单谱带（资料未显示）。

在泳道 6 和泳道 10 完全没有 220KD 的谱带是由于来自 MDTM 上清的无效沉淀是不可能的，因为在 S^{35} - 标记的 GH3Δ19 对照泳道（ 2 ）容易观察到 220KD 的谱带。利用含有野生型剪接位点的实施例 1 的 pDTM 构建体进行另一个分析结果显示在 350 和 220KD 处有两个强谱带。因此这些结果证明剪接位点的缺失导致产生没有 gp220 蛋白的 gp350 蛋白。

对在 CHO 细胞系，或其它的哺乳动物或非哺乳动物细胞系中表达的均一 gp350 蛋白质进一步测量，利用本领域内已知的方法包括例如凝集素亲和层析，反相 HPLC， FPLC，凝胶过滤等等技术从细胞系的限制性培养基分离或纯化均一 gp350 蛋白质。参见 David, J.Immunol.Methods 108:231(1988)和 Madej,Vaccine 10:777(1992)。

4.pMDTM 蛋白质的 Northern Blot 分析法

在该试验中我们采用 Northern Blot 分析法证明 MDTM - 1 细胞制备 gp350 RNA 和不制备 gp220 RNA，在另一个水平上证实剪接位点突变阻碍了 gp220 产生。

从 pDTM 制备与 gp350 互补的 DNA 探针，参见实施例 1 。分离 464 bp XhoI/PstI pDTM 片段形式的 gp350/220 探针模板 XP464 。 XP464 识别 gp350 和 gp220 。为了制备 gp350 特异性探针 AN537，用 NcoI 和 NdeI 切割 pDTM,分离两个重叠的 580bp 片段，用 XmnI 切割该混合物以去除污染片段并且用 AluI 切割产生位于 gp350/220 剪接位点内部的一

个 537 bp Alu/NdeI 片段。 AN537 特异于 gp220 之外的剪接区，因而特异于 gp350 信使。利用 DNA 片段 XP464 和 AN537 采用 ^{32}P - dCTP 缺口转译（ Amersham ）标记 DNA 探针。

基本上按照 Chomczyunski 和 Sacchi,Anal.Biochem.,162:156-59(1987 ）的方法制备总细胞 RNA 。从 CHO - pEE14（阴性对照细胞系，参见实施例 1 ）， CHO - MDTM - 1 和 CHO - DTM - 7 细胞，90 % 铺满，各两个 T - 250 培养瓶吸取培养基，裂解细胞并且刮入到变性缓冲液（ 10ml 4M 硫氰酸胍， 25mM 柠檬酸钠， pH7 ， 0.5%sarkosyl,100mM 2-巯基乙醇）。 10ml 裂解液中各补充 1ml 2M 乙酸钠， pH4 ,10ml 饱和酚， pH4.5 ,和 2 ml 氯仿/异戊醇；将裂解液在冰上放置 15 分钟，在 4 ℃以 10000 × g 旋转 20 分钟，除去上层水相。通过加入 1 倍体积的异丙醇在 4 ℃从水相沉淀 RNA 1 小时，在 4 ℃形成沉淀，重悬于变性缓冲液并且再沉淀。在 70 % 乙醇中将 RNA 洗涤 1 次，在 Speed-Vac 中干燥并且重悬于 DEPC 处理的水中。

在 15 % 甲酰胺和 6 % 甲醛中， 65 ℃将 DTM - 7 和 MDTM - 1 总细胞 RNA 变性 15 分钟，在 1 % 琼脂糖/6.6 % 甲醛凝胶上电泳，并且通过毛细作用转移到硝酸纤维素膜上并且用标记的 XP464 和 AN537 探测。通过煮沸 5 分钟使 DNA 探针变性，在 65 。 C 的 5 × SSPE 中杂交过夜，在高严谨条件下洗涤硝酸纤维素膜。利用 Bio-Rad 磷光体成象仪 (phospho image)进行放射自显影。

来自 CHO-MTDM 细胞的总细胞 RNA 的 Northern 印迹显示剪接突变在阻碍 gp220 特异性 RNA 产生中的效果。正如对于 gp350 特异性探针预期的， gp350 特异性探针， AN537 仅结合到 MDTM-1 和 DTM - 7 细胞中的一种类型的 RNA 上(附图 4 ，泳道 1 和 2)。特异于 gp350 和 gp220 RNA 的 XP464 探针识别 DTM - 7 的两个类型和 MTDM - 1 的单个较高分子量谱带(附图 4 ，泳道 4 和 3)，正如所预期的，如果剪接突变阻碍， MTDM 中产生 gp220 信使。即使对于 gp350 特异性 RNA ， MTDM 泳道过载，没有清晰的 gp220 RNA 观察到。 MTDM-1 和 DTM - 7 泳道中 gp350 信使的表观移动有差异的原因尚不知。与 DTM - 7 相比， MTDM - 7 类型过载，影响其在凝胶上的迁移。在凝胶上 gp350 信使迁移在靠近大核糖体 RNA 谱带处，这可能影响了表观分子量。或者存在与 gp350DNA 序列互补的单一类型的情况表明该信号

代表gp350 RNA。因此,供体和受体剪接位点中的突变有效地阻碍gp220信使的产生，正如 Northern 印迹判断的。该结果进一步由利用特异于gp350/220的单克隆抗体进行的放射自显影证实，也可以以 MDTM － 1和 DTM － 7 上清液的 Western 印迹分析证实。

实施例 4

检测均一 gp350 蛋白质的免疫原活性

将纯化的均一gp350蛋白质掺入到用于给药的合适的赋形剂中并且按照如下所述给药。

根据 David， J.Immunol.Methods 108:231(1988)和 Allison，J.Immunol.Methods 95:175(1986)的方法，将溶于 0.4%(v/v)吐温80Pluronic L121 和角鲨烯和(Thr^1)MDP 在磷酸缓冲盐水中混合制备 2×助剂－赋形剂溶液。

在给药当天通过加入等体积的蛋白质和助剂－赋形剂制备给药组合物。其蛋白质含量应该在 5mg-50mg/剂量范围内。

在 0， 21 和 42 天用 3 次 0.1 ml 肌内注射免疫接种 BALB/c。每次注射 10 天之后从 retro-orbital sinus 获取预免疫接种的血或连续取血。

根据实施例3描述的方法采用ELISA测定血清抗体水平。根据Moss, J Gen. Virol.17:233(1972)和 De Schryver, Int. J.Cancer 13:353(1974)的方法由其体外抑制 EBV 转化胎心脏淋巴细胞的能力对血清中的 EBV 中和抗体进行定量测定。

另一种方法，在 0， 21， 42， 63 和 84 天给新西兰白兔肌内注射接种 5 剂量的在前面所述助剂中乳化的蛋白质。每次接种的剂量应该在约 5μg-50μg 的范围内。末次剂量注射之后两星期取血清，检测抗体对该抗原的效价，检测对来自 B95－8 细胞的病毒 gp350/220 的交叉反应抗体和根据 Emini.Virology 166:387(1988)的方法检测体外 EBV 中和活性。

由于已经确立EBV gp350/220蛋白质能够在EBV感染的动物模型中诱导保护性免疫作用，参见 Epstein,Clin,Exp.Immunol 63:485(1986)，预期通过以均一 gp350 蛋白质组合物给药可获得类似的阳性结果。

本文中应用的所有出版物引入本文作参考。前面的详细描述是为了更清楚地了解，不应该认为限制本发明，因为对本领域内技术人员而言，本发明内的修改是显而易见的。

序列表

（1）一般资料:

（i）申请人: Spaete,Richard and Jackman,Winthrop,T.

（ii）发明名称: gp350/220 的非剪接变异体

（iii）序列数: 19

（iv）相应地址:

（A）收件人: Cooley Godward Castro Huddleson & Tatum

（B）街道: 5 Palo Alto Square

（C）城市: Palo Alto

（D）州: California

（E）国家: USA

（F）ZIP: 94306

（v）计算机可读形式

（A）媒体类型: Floppy disk

（B）计算机: IBM PC Compatible

（C）操作系统: PC - DOS/MS - DOS

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

（vi）现在申请资料

（A）申请号: 08/229,291

（B）申请日: April 18,1994

（C）分类:

（Viii）代理人/律师资料

（A）姓名: Luann Cserr

（B）登记号: 31,822

（C）档案号: AVIR-003/00US

（ix）电讯资料

（A）电话: 415 - 843 - 5163

（B）传真: 415 - 857 - 0663

（C）电传: 380816 CooleyPA

（2）SEQ ID NO: 1 的资料

（i）序列特征:

（A）长度: 27 碱基对

（B）类型：核酸

（C）链数：单

（D）拓朴：未知

（ii）分子类型：寡聚体 DNA

（xi）序列描述：SEQ ID NO：1

GGATCCTAGA CTGCGCCTTT AGGCGTA

（2）SEQ ID NO：2 的资料

（i）序列特征：

（A）长度：19 碱基对

（B）类型：核酸

（C）链数：单

（D）拓朴：未知

（ii）分子类型：寡聚体 DNA

（xi）SEQ ID NO：2 序列描述

GACTGCGCCT TTAGGCGTA

（2）SEQ ID NO：3 的资料

（i）序列特征：

（A）长度：6 氨基酸

（B）类型：氨基酸

（C）拓朴：线性

（ii）分子类型：蛋白质

（xi）序列描述：SEQ ID NO：3

Asp Cys Ala Phe Arg Arg

1 5

（2）SEQ ID NO：4 的资料

（i）序列特征：

（A）长度：27 个碱基

（B）类型：核酸

（C）链：单

（D）拓朴：未知

（ii）分子类型：寡聚体 DNA

（xi）序列描述：SEQ ID NO：4

GGATCCTCTG TTCCTTCTGC TCCAGTG

（2）SEQ ID NO：5 的资料

（i）序列特征:

（A）长度：21 个碱基

（B）类型：核酸

（C）链：单

（D）拓扑：未知

（ii）分子类型：寡聚体 DNA

（xi）序列描述：SEQ ID NO：5

TCTGTTCCTT CTGCTCCAGT G

（2）SEQ ID NO：6 的资料

（i）序列特征:

（A）长度：16 个碱基

（B）类型：核酸

（C）链：单

（D）拓扑：未知

（ii）分子类型：寡聚体 DNA

（xi）序列描述：SEQ ID NO：6

TATAGACTAG TCTAGG

（2）SEQ ID NO：7 的资料

（i）序列特征:

（A）长度：18 个碱基

（B）类型：核酸

（C）链：单

（D）拓扑：未知

（ii）分子类型：寡聚体 DNA

（xi）序列描述：SEQ ID NO：7

ATCTGATCAG ATCCTTAA

（2）SEQ ID NO：8 的资料

（ i ）序列特征:

（ A ）长度: 39 个碱基

（ B ）类型: 核酸

（ C ）链: 单

（ D ）拓扑: 未知

（ ii ）分子类型: 寡聚体 DNA

（ xi ）序列描述: SEQ ID NO: 8

AACCTCTCCA TGCTAGTACT GGTCATGGCG GACTGCGCC

（ 2 ）SEQ ID NO: 9 的资料

（ i ）序列特征:

（ A ）长度: 13 氨基酸

（ B ）类型: 氨基酸

（ C ）拓扑: 线性

（ ii ）分子类型: 蛋白质

（ xi ）序列描述: SEQ ID NO: 9

Asn Leu Ser Met Leu Val Leu Val Met Ala Asp Cys Ala

1 5 10

（ 2 ）SEQ ID NO: 10 的资料

（ i ）序列特征:

（ A ）长度: 30 个碱基

（ B ）类型: 核酸

（ C ）链: 单

（ D ）拓扑: 未知

（ ii ）分子类型: 寡聚体 DNA

（ xi ）序列描述: SEQ ID NO: 10

AACCTCTCCA TGCTATAGAC TAGTTCTAGG

（ 2 ）SEQ ID NO: 11 的资料

（ i ）序列特征:

（ A ）长度: 5 个氨基酸

（B）类型：氨基酸
（C）拓扑：线性
（ii）分子类型：蛋白质
（xi）序列描述：SEQ ID NO：11

Asn Leu Ser Met Leu
1 5

（2）SEQ ID NO：12 的资料
（i）序列特征：
（A）长度：24 个碱基
（B）类型：核酸
（C）链：单
（D）拓扑：未知
（ii）分子类型：寡聚体 DNA
（xi）序列描述：SEQ ID NO：12

AACCTCTCCA TGCTAGACTG CGCC

（2）SEQ ID NO：13 的资料
（i）序列特征：
（A）长度：8 个氨基酸
（B）类型：氨基酸
（C）拓扑：线性
（ii）分子类型：蛋白质
（xi）序列描述：SEQ ID NO：13

Asn Leu Ser Met Leu_{862} Asp_{882} Cys Ala
1 5

（2）SEQ ID NO：14 的资料
（i）序列特征：
（A）长度：42 个碱基
（B）类型：核酸
（C）链：单

（ D ）拓扑：未知

（ ii ）分子类型：寡聚体 DNA

（ xi ）序列描述： SEQ ID NO： 14

GGTCATGTCG GGGGCCTTTG ACTCTGTGCC GTTGTCCCAT GG

（ 2 ） SEQ ID NO： 15 的资料

（ i ）序列特征：

（ A ）长度： 42 个碱基

（ B ）类型：核酸

（ C ）链：单

（ D ）拓扑：未知

（ ii ）分子类型：寡聚体 DNA

（ xi ）序列描述： SEQ ID NO： 15

GGTCATGTCG GGGGCCTTAC TTTCTGTGCC GTTGTCCCAT GG

（ 2 ） SEQ ID NO： 16 的资料

（ i ）序列特征：

（ A ）长度： 42 个碱基

（ B ）类型：核酸

（ C ）链：单

（ D ）拓扑：未知

（ ii ）分子类型：寡聚体 DNA

（ xi ）序列描述： SEQ ID NO： 16

CTGTGTTATA TTTTCACCTC CAGTTGGGTG AGCGGAGGTT AG

（ 2 ） SEQ ID NO： 18 信息

（ i ）序列特征：

（ A ）长度： 3833 碱基

（ B ）类型：核酸

（ C ）链型：双链

（ D ）拓扑：未知

（ ii ）分子类型： cDNA

（ xi ）特点：

（ A ）名称/键： CDS

（ B ）位置： 1014..3734

(xi) 序列描述

GAATTCCATA AATGAAACAC GCTGGTCAGG TGTTAAAACT TCCTCCCAGA TTTTCGTGAG 60

GCTCCTGTGT ATAGCCATAT AGTCAAAGAA AATACTGTAG CGGGGATTAC AGCTCTGTAC 120

AATGTTACCC ACGGAGCTCT GAACATACAA CCACTGGCGA TCCCCGGGGG TACATCGCGG 180

CAGCTTAAAG GTGCCGGCGG AAAAGGTCAC GTGACACCTA CGGCCACCTG TGCACCCAAG 240

TGTCGCCTGG AGATGTACGA ATGTGGGAGT CGTCTGGTGA TCGGTGTAGC TGTACATCCA 300

GCTGCTGTAT GCCTGGTAAC CCATAGGCCA TCCGGCGGCC AGGGTTTGCA GTCTCCATTT 360

GGCCTGATCT CTACGAGAAG CTGGATTTCT CCGACGATCT CTAATGGCCT GTCGAATGGC 420

CATGGCATAC ATTATGTACA TCTCGGTATT TGAAATCTGG ATCCGAAAAA CTGGTCTATG 480

GCTCGTGTGT CGATGCGCTG AAACCAACGG CAACAAATTA CTTACCTTGT TGTTGTGTGA 540

TGGGTAAAAA CACACATCAC ACACTTAGGC CATAGGGATG CTCACCGTAG CCGCGGCTCC 600

AATCGCTTGA AGAAGTGTTC TTAGATCTAG TGGAAACCTG CGGAGAATGG CTTCTCGCCC 660

AGGGAGATCC GGCTGGGGTG GGAGCATGGG TCGTGCTGGA GCTGACCCAC CGGCATCATG 720

ATCGACCCGC TTTCTCTTCG TACCCTTCTG GGCCGGCTCC AGGTGGGCAT CTTCTGCTTC 780

CTTTTCTGAG CTGCTATCTG ATAACTCTAT GAGGACATTT TCCCAATCTC CCGCCGATAC 840

CTGTTCCTGC ACAACCGAGG TAGATGGGAC TTCTTCTTCC ATGTTGTCAT CCAGGGCCGG 900

GGGACCCGGC CTGTCCTTGT CCATTTTGTC TGCAACAAAA GTGTGACTCA CCAACACCGC 960

ACCCCCCTTG TACCTATTAA AGAGGATGCT GCCTAGAAAT CGGTGCCGAG ACA ATG 1016
Met
1

GAG GCA GCC TTG CTT GTG TGT CAG TAC ACC ATC CAG AGC CTG ATC CAT 1064
Glu Ala Ala Leu Leu Val Cys Gln Tyr Thr Ile Gln Ser Leu Ile His
5 10 15

```
CTC ACG GGT GAA GAT CCT GGT TTT TTC AAT GTT GAG ATT CCG GAA TTC      1112
Leu Thr Gly Glu Asp Pro Gly Phe Phe Asn Val Glu Ile Pro Glu Phe
          20                  25                  30

CCA TTT TAC CCC ACA TGC AAT GTT TGC ACG GCA GAT GTC AAT GTA ACT      1160
Pro Phe Tyr Pro Thr Cys Asn Val Cys Thr Ala Asp Val Asn Val Thr
      35                  40                  45

ATC AAT TTC GAT GTC GGG GGC AAA AAG CAT CAA CTT GAT CTT GAC TTT      1208
Ile Asn Phe Asp Val Gly Gly Lys Lys His Gln Leu Asp Leu Asp Phe
 50                  55                  60                  65

GGC CAG CTG ACA CCC CAT ACG AAG GCT GTC TAC CAA CCT CGA GGT GCA      1256
Gly Gln Leu Thr Pro His Thr Lys Ala Val Tyr Gln Pro Arg Gly Ala
                 70                  75                  80

TTT GGT GGC TCA GAA AAT GCC ACC AAT CTC TTT CTA CTG GAG CTC CTT      1304
Phe Gly Gly Ser Glu Asn Ala Thr Asn Leu Phe Leu Leu Glu Leu Leu
             85                  90                  95

GGT GCA GGA GAA TTG GCT CTA ACT ATG CGG TCT AAG AAG CTT CCA ATT      1352
Gly Ala Gly Glu Leu Ala Leu Thr Met Arg Ser Lys Lys Leu Pro Ile
         100                 105                 110

AAC GTC ACC ACC GGA GAG GAG CAA CAA GTA AGC CTG GAA TCT GTA GAT      1400
Asn Val Thr Thr Gly Glu Glu Gln Gln Val Ser Leu Glu Ser Val Asp
     115                 120                 125

GTC TAC TTT CAA GAT GTG TTT GGA ACC ATG TGG TGC CAC CAT GCA GAA      1448
Val Tyr Phe Gln Asp Val Phe Gly Thr Met Trp Cys His His Ala Glu
 130                 135                 140                 145

ATG CAA AAC CCC GTG TAC CTG ATA CCA GAA ACA GTG CCA TAC ATA AAG      1496
Met Gln Asn Pro Val Tyr Leu Ile Pro Glu Thr Val Pro Tyr Ile Lys
                150                 155                 160

TGG GAT AAC TGT AAT TCT ACC AAT ATA ACG GCA GTA GTG AGG GCA CAG      1544
Trp Asp Asn Cys Asn Ser Thr Asn Ile Thr Ala Val Val Arg Ala Gln
             165                 170                 175

GGG CTG GAT GTC ACG CTA CCC TTA AGT TTG CCA ACG TCA GCT CAA GAC      1592
Gly Leu Asp Val Thr Leu Pro Leu Ser Leu Pro Thr Ser Ala Gln Asp
         180                 185                 190
```

```
TCG AAT TTC AGC GTA AAA ACA GAA ATG CTC GGT AAT GAG ATA GAT ATT      1640
Ser Asn Phe Ser Val Lys Thr Glu Met Leu Gly Asn Glu Ile Asp Ile
    195                 200                 205

GAG TGT ATT ATG GAG GAT GGC GAA ATT TCA CAA GTT CTG CCC GGA GAC      1688
Glu Cys Ile Met Glu Asp Gly Glu Ile Ser Gln Val Leu Pro Gly Asp
210                 215                 220                 225

AAC AAA TTT AAC ATC ACC TGC AGT GGA TAC GAG AGC CAT GTT CCC AGC      1736
Asn Lys Phe Asn Ile Thr Cys Ser Gly Tyr Glu Ser His Val Pro Ser
                230                 235                 240

GGC GGA ATT CTC ACA TCA ACG AGT CCC GTG GCC ACC CCA ATA CCT GGT      1784
Gly Gly Ile Leu Thr Ser Thr Ser Pro Val Ala Thr Pro Ile Pro Gly
            245                 250                 255

ACA GGG TAT GCA TAC AGC CTG CGT CTG ACA CCA CGT CCA GTG TCA CGA      1832
Thr Gly Tyr Ala Tyr Ser Leu Arg Leu Thr Pro Arg Pro Val Ser Arg
        260                 265                 270

TTT CTT GGC AAT AAC AGT ATC CTG TAC GTG TTT TAC TCT GGG AAT GGA      1880
Phe Leu Gly Asn Asn Ser Ile Leu Tyr Val Phe Tyr Ser Gly Asn Gly
    275                 280                 285

CCG AAG GCG AGC GGG GGA GAT TAC TGC ATT CAG TCC AAC ATT GTG TTC      1928
Pro Lys Ala Ser Gly Gly Asp Tyr Cys Ile Gln Ser Asn Ile Val Phe
290                 295                 300                 305

TCT GAT GAG ATT CCA GCT TCA CAG GAC ATG CCG ACA AAC ACC ACA GAC      1976
Ser Asp Glu Ile Pro Ala Ser Gln Asp Met Pro Thr Asn Thr Thr Asp
                310                 315                 320

ATC ACA TAT GTG GGT GAC AAT GCT ACC TAT TCA GTG CCA ATG GTC ACT      2024
Ile Thr Tyr Val Gly Asp Asn Ala Thr Tyr Ser Val Pro Met Val Thr
            325                 330                 335

TCT GAG GAC GCA AAC TCG CCA AAT GTT ACA GTG ACT GCC TTT TGG GCC      2072
Ser Glu Asp Ala Asn Ser Pro Asn Val Thr Val Thr Ala Phe Trp Ala
        340                 345                 350

TGG CCA AAC AAC ACT GAA ACT GAC TTT AAG TGC AAA TGG ACT CTC ACC      2120
Trp Pro Asn Asn Thr Glu Thr Asp Phe Lys Cys Lys Trp Thr Leu Thr
    355                 360                 365
```

```
TCG GGG ACA CCT TCG GGT TGT GAA AAT ATT TCT GGT GCA TTT GCG AGC      2168
Ser Gly Thr Pro Ser Gly Cys Glu Asn Ile Ser Gly Ala Phe Ala Ser
370                 375                 380                 385

AAT CGG ACA TTT GAC ATT ACT GTC TCG GGT CTT GGC ACG GCC CCC AAG      2216
Asn Arg Thr Phe Asp Ile Thr Val Ser Gly Leu Gly Thr Ala Pro Lys
                390                 395                 400

ACA CTC ATT ATC ACA CGA ACG GCT ACC AAT GCC ACC ACA ACA ACC CAC      2264
Thr Leu Ile Ile Thr Arg Thr Ala Thr Asn Ala Thr Thr Thr Thr His
            405                 410                 415

AAG GTT ATA TTC TCC AAG GCA CCC GAG AGC ACC ACC ACC TCC CCT ACC      2312
Lys Val Ile Phe Ser Lys Ala Pro Glu Ser Thr Thr Thr Ser Pro Thr
        420                 425                 430

TTG AAT ACA ACT GGA TTT GCT GAT CCC AAT ACA ACG ACA GGT CTA CCC      2360
Leu Asn Thr Thr Gly Phe Ala Asp Pro Asn Thr Thr Thr Gly Leu Pro
    435                 440                 445

AGC TCT ACT CAC GTG CCT ACC AAC CTC ACC GCA CCT GCA AGC ACA GGC      2408
Ser Ser Thr His Val Pro Thr Asn Leu Thr Ala Pro Ala Ser Thr Gly
450                 455                 460                 465

CCC ACT GTA TCC ACC GCG GAT GTC ACC AGC CCA ACA CCA GCC GGC ACA      2456
Pro Thr Val Ser Thr Ala Asp Val Thr Ser Pro Thr Pro Ala Gly Thr
                470                 475                 480

ACG TCA GGC GCA TCA CCG GTG ACA CCA AGT CCA TCT CCA TGG GAC AAC      2504
Thr Ser Gly Ala Ser Pro Val Thr Pro Ser Pro Ser Pro Trp Asp Asn
            485                 490                 495

GGC ACA GAA AGT AAG GCC CCC GAC ATG ACC AGC TCC ACC TCA CCA GTG      2552
Gly Thr Glu Ser Lys Ala Pro Asp Met Thr Ser Ser Thr Ser Pro Val
        500                 505                 510

ACT ACC CCA ACC CCA AAT GCC ACC AGC CCC ACC CCA GCA GTG ACT ACC      2600
Thr Thr Pro Thr Pro Asn Ala Thr Ser Pro Thr Pro Ala Val Thr Thr
    515                 520                 525

CCA ACC CCA AAT GCC ACC AGC CCC ACC CCA GCA GTG ACT ACC CCA ACC      2648
Pro Thr Pro Asn Ala Thr Ser Pro Thr Pro Ala Val Thr Thr Pro Thr
530                 535                 540                 545
```

```
CCA AAT GCC ACC AGC CCC ACC TTG GGA AAA ACA AGT CCT ACC TCA GCA      2696
Pro Asn Ala Thr Ser Pro Thr Leu Gly Lys Thr Ser Pro Thr Ser Ala
                550                 555                 560

GTG ACT ACC CCA ACC CCA AAT GCC ACC AGC CCC ACC TTG GGA AAA ACA      2744
Val Thr Thr Pro Thr Pro Asn Ala Thr Ser Pro Thr Leu Gly Lys Thr
            565                 570                 575

AGC CCC ACC TCA GCA GTG ACT ACC CCA ACC CCA AAT GCC ACC AGC CCC      2792
Ser Pro Thr Ser Ala Val Thr Thr Pro Thr Pro Asn Ala Thr Ser Pro
        580                 585                 590

ACC TTG GGA AAA ACA AGC CCC ACC TCA GCA GTG ACT ACC CCA ACC CCA      2840
Thr Leu Gly Lys Thr Ser Pro Thr Ser Ala Val Thr Thr Pro Thr Pro
    595                 600                 605

AAT GCC ACC GGC CCT ACT GTG GGA GAA ACA AGT CCA CAG GCA AAT GCC      2888
Asn Ala Thr Gly Pro Thr Val Gly Glu Thr Ser Pro Gln Ala Asn Ala
610                 615                 620                 625

ACC AAC CAC ACC TTA GGA GGA ACA AGT CCC ACC CCA GTA GTT ACC AGC      2936
Thr Asn His Thr Leu Gly Gly Thr Ser Pro Thr Pro Val Val Thr Ser
                630                 635                 640

CAA CCA AAA AAT GCA ACC AGT GCT GTT ACC ACA GGC CAA CAT AAC ATA      2984
Gln Pro Lys Asn Ala Thr Ser Ala Val Thr Thr Gly Gln His Asn Ile
            645                 650                 655

ACT TCA AGT TCA ACC TCT TCC ATG TCA CTG AGA CCC AGT TCA AAC CCA      3032
Thr Ser Ser Ser Thr Ser Ser Met Ser Leu Arg Pro Ser Ser Asn Pro
        660                 665                 670

GAG ACA CTC AGC CCC TCC ACC AGT GAC AAT TCA ACG TCA CAT ATG CCT      3080
Glu Thr Leu Ser Pro Ser Thr Ser Asp Asn Ser Thr Ser His Met Pro
    675                 680                 685

TTA CTA ACC TCC GCT CAC CCA AGA GGT GGT GAA AAT ATA ACA CAG GTG      3128
Leu Leu Thr Ser Ala His Pro Thr Gly Gly Glu Asn Ile Thr Gln Val
690                 695                 700                 705

ACA CCA GCC TCT ATC AGC ACA CAT CAT GTG TCC ACC AGT TCG CCA GAA      3176
Thr Pro Ala Ser Ile Ser Thr His His Val Ser Thr Ser Ser Pro Glu
                710                 715                 720
```

```
   CCC CGC CCA GGC ACC ACC AGC CAA GCG TCA GGC CCT GGA AAC AGT TCC      3224
   Pro Arg Pro Gly Thr Thr Ser Gln Ala Ser Gly Pro Gly Asn Ser Ser
               725                 730                 735

 5 ACA TCC ACA AAA CCG GGG GAG GTT AAT GTC ACC AAA GGC ACG CCC CCC      3272
   Thr Ser Thr Lys Pro Gly Glu Val Asn Val Thr Lys Gly Thr Pro Pro
           740                 745                 750

   CAA AAT GCA ACG TCG CCC CAG GCC CCC AGT GGC CAA AAG ACG GCG GTT      3320
10 Gln Asn Ala Thr Ser Pro Gln Ala Pro Ser Gly Gln Lys Thr Ala Val
       755                 760                 765

   CCC ACG GTC ACC TCA ACA GGT GGA AAG GCC AAT TCT ACC ACC GGT GGA      3368
   Pro Thr Val Thr Ser Thr Gly Gly Lys Ala Asn Ser Thr Thr Gly Gly
15 770                 775                 780                 785

   AAG CAC ACC ACA GGA CAT GGA GCC CGG ACA AGT ACA GAG CCC ACC ACA      3416
   Lys His Thr Thr Gly His Gly Ala Arg Thr Ser Thr Glu Pro Thr Thr
                   790                 795                 800
20
   GAT TAC GGC GGT GAT TCA ACT ACG CCA AGA CCG AGA TAC AAT GCG ACC      3464
   Asp Tyr Gly Gly Asp Ser Thr Thr Pro Arg Pro Arg Tyr Asn Ala Thr
               805                 810                 815

25 ACC TAT CTA CCT CCC AGC ACT TCT AGC AAA CTG CGG CCC CGC TGG ACT      3512
   Thr Tyr Leu Pro Pro Ser Thr Ser Ser Lys Leu Arg Pro Arg Trp Thr
           820                 825                 830

   TTT ACG AGC CCA CCG GTT ACC ACA GCC CAA GCC ACC GTG CCA GTC CCG      3560
30 Phe Thr Ser Pro Pro Val Thr Thr Ala Gln Ala Thr Val Pro Val Pro
       835                 840                 845

   CCA ACG TCC CAG CCC AGA TTC TCA AAC CTC TCC ATG CTA GTA CTG CAG      3608
   Pro Thr Ser Gln Pro Arg Phe Ser Asn Leu Ser Met Leu Val Leu Gln
35 850                 855                 860                 865

   TGG GCC TCT CTG GCT GTG CTG ACC CTT CTG CTG CTG CTG GTC ATG GCG      3656
   Trp Ala Ser Leu Ala Val Leu Thr Leu Leu Leu Leu Leu Val Met Ala
                   870                 875                 880
40
   GAC TGC GCC TTT AGG CGT AAC TTG TCT ACA TCC CAT ACC TAC ACC ACC      3704
   Asp Cys Ala Phe Arg Arg Asn Leu Ser Thr Ser His Thr Tyr Thr Thr
               885                 890                 895
```

```
CCA CCA TAT GAT GAC GCC GAG ACC TAT GTA TAAAGTCAAT AAAAATTTAT        3754
Pro Pro Tyr Asp Asp Ala Glu Thr Tyr Val
        900                 905

TAATCAGAAA TTTGCACTTT CTTTGCTTCA CGTCCCCGGG AGCGGGAGCG GGCACGTCGG   3814

GTGGCGTTGG GGTCGTTTG                                                3833
```

（2）SEQ ID NO：19

（i）序列特征

（A）长度：907 氨基酸

（B）类型：氨基酸

（C）链：单

（ii）分子类型：蛋白质

（xi）序列描述：SEQ ID NO：19

```
Met Glu Ala Ala Leu Leu Val Cys Gln Tyr Thr Ile Gln Ser Leu Ile
 1               5                  10                  15

His Leu Thr Gly Glu Asp Pro Gly Phe Phe Asn Val Glu Ile Pro Glu
            20                  25                  30

Phe Pro Phe Tyr Pro Thr Cys Asn Val Cys Thr Ala Asp Val Asn Val
        35                  40                  45

Thr Ile Asn Phe Asp Val Gly Gly Lys Lys His Gln Leu Asp Leu Asp
    50                  55                  60

Phe Gly Gln Leu Thr Pro His Thr Lys Ala Val Tyr Gln Pro Arg Gly
65                  70                  75                  80

Ala Phe Gly Gly Ser Glu Asn Ala Thr Asn Leu Phe Leu Leu Glu Leu
                85                  90                  95

Leu Gly Ala Gly Glu Leu Ala Leu Thr Met Arg Ser Lys Lys Leu Pro
            100                 105                 110

Ile Asn Val Thr Thr Gly Glu Glu Gln Gln Val Ser Leu Glu Ser Val
        115                 120                 125
```

```
Asp Val Tyr Phe Gln Asp Val Phe Gly Thr Met Trp Cys His His Ala
    130                 135                 140

Glu Met Gln Asn Pro Val Tyr Leu Ile Pro Glu Thr Val Pro Tyr Ile
145                 150                 155                 160

Lys Trp Asp Asn Cys Asn Ser Thr Asn Ile Thr Ala Val Val Arg Ala
                165                 170                 175

Gln Gly Leu Asp Val Thr Leu Pro Leu Ser Leu Pro Thr Ser Ala Gln
            180                 185                 190

Asp Ser Asn Phe Ser Val Lys Thr Glu Met Leu Gly Asn Glu Ile Asp
        195                 200                 205

Ile Glu Cys Ile Met Glu Asp Gly Glu Ile Ser Gln Val Leu Pro Gly
    210                 215                 220

Asp Asn Lys Phe Asn Ile Thr Cys Ser Gly Tyr Glu Ser His Val Pro
225                 230                 235                 240

Ser Gly Gly Ile Leu Thr Ser Thr Ser Pro Val Ala Thr Pro Ile Pro
                245                 250                 255

Gly Thr Gly Tyr Ala Tyr Ser Leu Arg Leu Thr Pro Arg Pro Val Ser
            260                 265                 270

Arg Phe Leu Gly Asn Asn Ser Ile Leu Tyr Val Phe Tyr Ser Gly Asn
        275                 280                 285

Gly Pro Lys Ala Ser Gly Gly Asp Tyr Cys Ile Gln Ser Asn Ile Val
    290                 295                 300

Phe Ser Asp Glu Ile Pro Ala Ser Gln Asp Met Pro Thr Asn Thr Thr
305                 310                 315                 320

Asp Ile Thr Tyr Val Gly Asp Asn Ala Thr Tyr Ser Val Pro Met Val
                325                 330                 335

Thr Ser Glu Asp Ala Asn Ser Pro Asn Val Thr Val Thr Ala Phe Trp
            340                 345                 350

Ala Trp Pro Asn Asn Thr Glu Thr Asp Phe Lys Cys Lys Trp Thr Leu
```

```
          355                 360                 365

   Thr Ser Gly Thr Pro Ser Gly Cys Glu Asn Ile Ser Gly Ala Phe Ala
       370                 375                 380
5
   Ser Asn Arg Thr Phe Asp Ile Thr Val Ser Gly Leu Gly Thr Ala Pro
   385                 390                 395                 400

   Lys Thr Leu Ile Ile Thr Arg Thr Ala Thr Asn Ala Thr Thr Thr Thr
10                 405                 410                 415

   His Lys Val Ile Phe Ser Lys Ala Pro Glu Ser Thr Thr Thr Ser Pro
              420                 425                 430

15 Thr Leu Asn Thr Thr Gly Phe Ala Asp Pro Asn Thr Thr Thr Gly Leu
          435                 440                 445

   Pro Ser Ser Thr His Val Pro Thr Asn Leu Thr Ala Pro Ala Ser Thr
       450                 455                 460
20
   Gly Pro Thr Val Ser Thr Ala Asp Val Thr Ser Pro Thr Pro Ala Gly
   465                 470                 475                 480

   Thr Thr Ser Gly Ala Ser Pro Val Thr Pro Ser Pro Ser Pro Trp Asp
25                 485                 490                 495

   Asn Gly Thr Glu Ser Lys Ala Pro Asp Met Thr Ser Ser Thr Ser Pro
              500                 505                 510

30 Val Thr Thr Pro Thr Pro Asn Ala Thr Ser Pro Thr Pro Ala Val Thr
          515                 520                 525

   Thr Pro Thr Pro Asn Ala Thr Ser Pro Thr Pro Ala Val Thr Thr Pro
       530                 535                 540
35
   Thr Pro Asn Ala Thr Ser Pro Thr Leu Gly Lys Thr Ser Pro Thr Ser
   545                 550                 555                 560

   Ala Val Thr Thr Pro Thr Pro Asn Ala Thr Ser Pro Thr Leu Gly Lys
40                 565                 570                 575

   Thr Ser Pro Thr Ser Ala Val Thr Thr Pro Thr Pro Asn Ala Thr Ser
              580                 585                 590
```

```
Pro Thr Leu Gly Lys Thr Ser Pro Thr Ser Ala Val Thr Thr Pro Thr
        595                 600                 605

Pro Asn Ala Thr Gly Pro Thr Val Gly Glu Thr Ser Pro Gln Ala Asn
    610                 615                 620

Ala Thr Asn His Thr Leu Gly Gly Thr Ser Pro Thr Pro Val Val Thr
625                 630                 635                 640

Ser Gln Pro Lys Asn Ala Thr Ser Ala Val Thr Thr Gly Gln His Asn
                645                 650                 655

Ile Thr Ser Ser Ser Thr Ser Ser Met Ser Leu Arg Pro Ser Ser Asn
            660                 665                 670

Pro Glu Thr Leu Ser Pro Ser Thr Ser Asp Asn Ser Thr Ser His Met
        675                 680                 685

Pro Leu Leu Thr Ser Ala His Pro Thr Gly Gly Glu Asn Ile Thr Gln
    690                 695                 700

Val Thr Pro Ala Ser Ile Ser Thr His His Val Ser Thr Ser Ser Pro
705                 710                 715                 720

Glu Pro Arg Pro Gly Thr Thr Ser Gln Ala Ser Gly Pro Gly Asn Ser
                725                 730                 735

Ser Thr Ser Thr Lys Pro Gly Glu Val Asn Val Thr Lys Gly Thr Pro
            740                 745                 750

Pro Gln Asn Ala Thr Ser Pro Gln Ala Pro Ser Gly Gln Lys Thr Ala
        755                 760                 765

Val Pro Thr Val Thr Ser Thr Gly Gly Lys Ala Asn Ser Thr Thr Gly
    770                 775                 780

Gly Lys His Thr Thr Gly His Gly Ala Arg Thr Ser Thr Glu Pro Thr
785                 790                 795                 800

Thr Asp Tyr Gly Gly Asp Ser Thr Thr Pro Arg Pro Arg Tyr Asn Ala
                805                 810                 815

Thr Thr Tyr Leu Pro Pro Ser Thr Ser Ser Lys Leu Arg Pro Arg Trp
```

```
                820                 825                 830

   Thr Phe Thr Ser Pro Pro Val Thr Thr Ala Gln Ala Thr Val Pro Val
           835                 840                 845

 5
   Pro Pro Thr Ser Gln Pro Arg Phe Ser Asn Leu Ser Met Leu Val Leu
       850                 855                 860

   Gln Trp Ala Ser Leu Ala Val Leu Thr Leu Leu Leu Leu Leu Val Met
10 865                 870                 875                 880

   Ala Asp Cys Ala Phe Arg Arg Asn Leu Ser Thr Ser His Thr Tyr Thr
                  885                 890                 895

15 Thr Pro Pro Tyr Asp Asp Ala Glu Thr Tyr Val
               900                 905
```

20

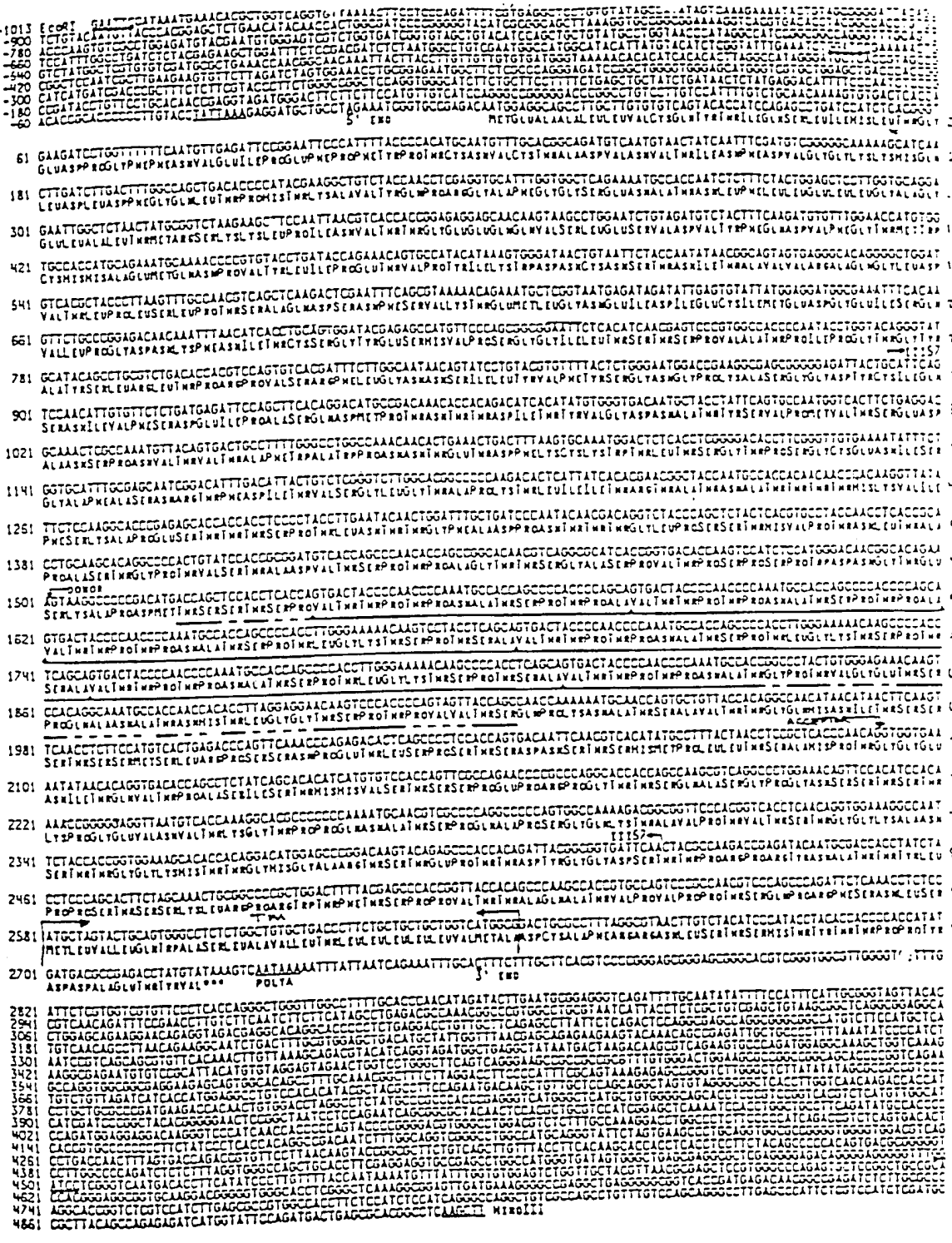


图 1

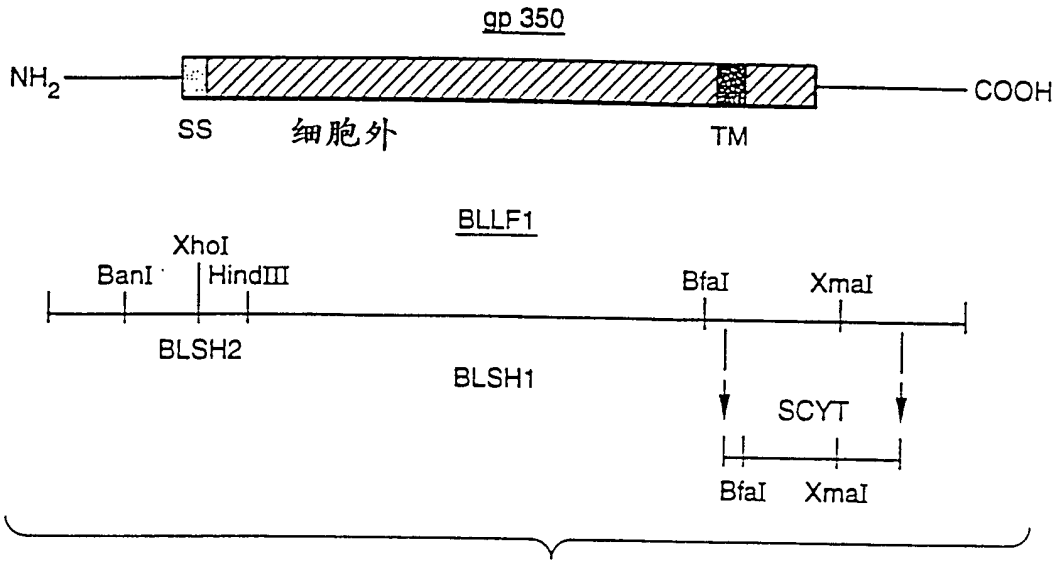

gp 350
NH2
COOH
SS
细胞外
TM
BLLF1
XhoI
BanI
HindIII
BfaI
XmaI
BLSH2
BLSH1
SCYT
BfaI
XmaI

图 2A

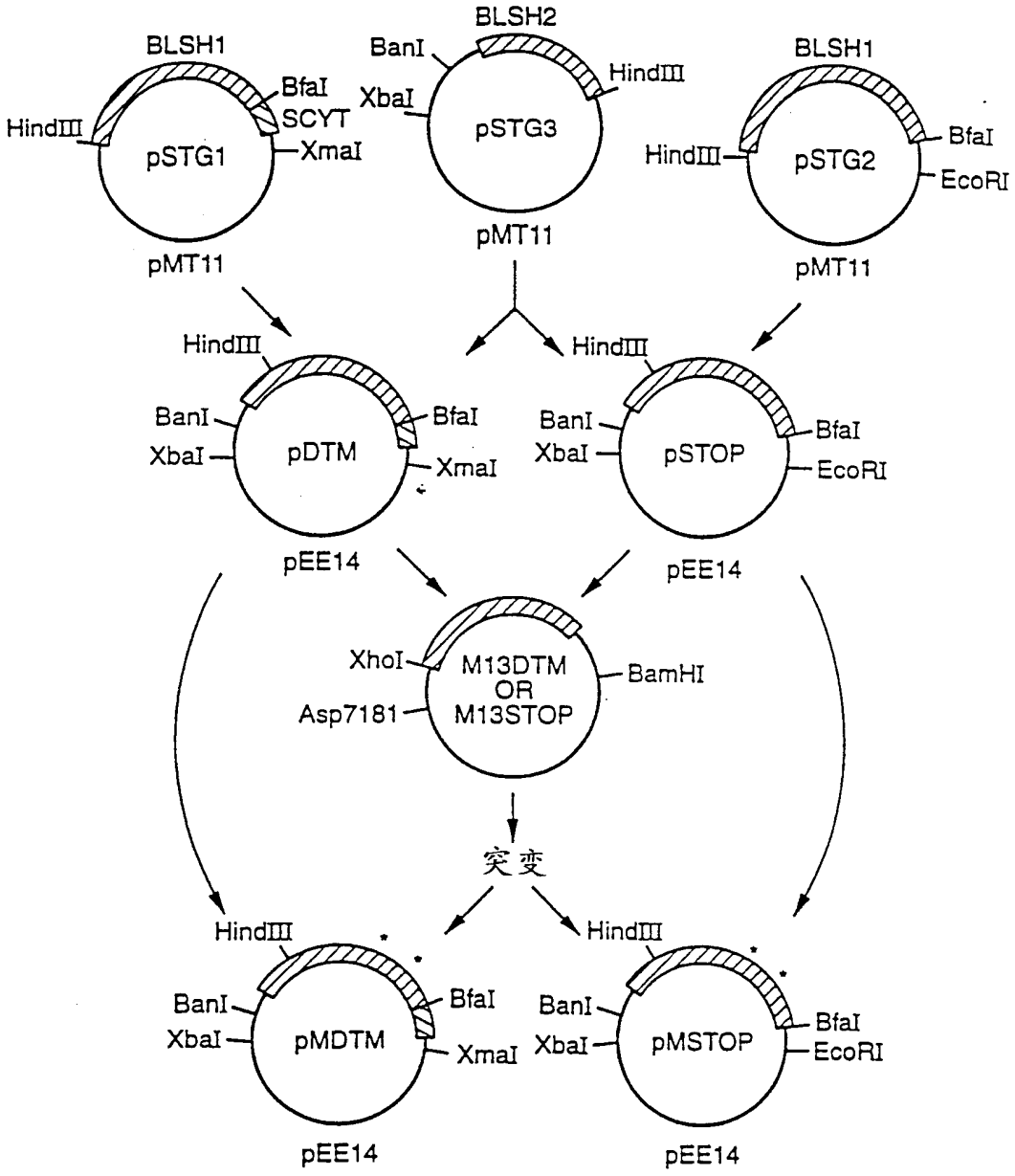

BLSH1
HindIII
BfaI
SCYT
XmaI
pSTG1
pMT11
BLSH2
BanI
XbaI
HindIII
pSTG3
pMT11
BLSH1
HindIII
BfaI
EcoRI
pSTG2
pMT11
HindIII
BanI
XbaI
BfaI
XmaI
pDTM
pEE14
HindIII
BanI
XbaI
BfaI
EcoRI
pSTOP
pEE14
XhoI
Asp7181
M13DTM
OR
M13STOP
BamHI
突变
HindIII
BanI
XbaI
BfaI
XmaI
pMDTM
pEE14
HindIII
BanI
XbaI
BfaI
EcoRI
pMSTOP
pEE14

图 2B

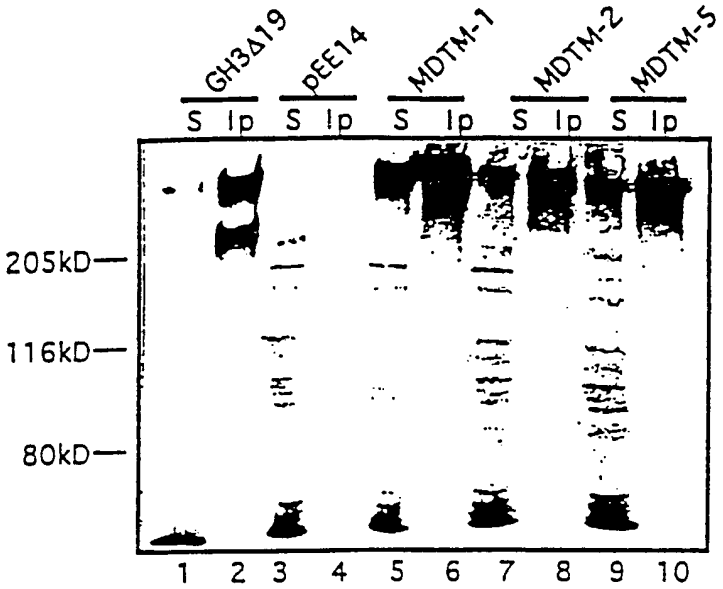

GH3Δ19
pEE14
MDTM-1
MDTM-2
MDTM-5
S Ip S Ip S Ip S Ip S Ip
205kD
116kD
80kD
1 2 3 4 5 6 7 8 9 10

图 3

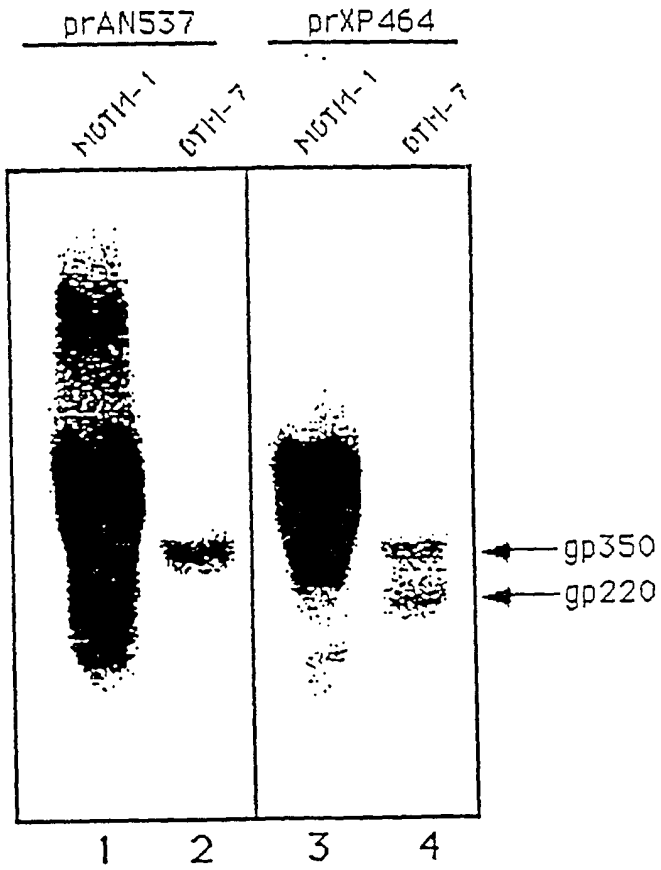


图 4