

1. 肺炎链球菌蛋白质或多肽，其序列选自表 1 所示序列。
2. 如权利要求 1 所要求的蛋白质或多肽，其以基本上纯的形式提供。
3. 与权利要求 1 或 2 定义的一种蛋白质或多肽基本相同的蛋白质或多肽。
4. 权利要求 1 到 3 任一项所定义的蛋白质或多肽的同源物或衍生物。
5. 为表 1 和 3 所定义的蛋白质或多肽的抗原性和/或免疫原性片段。
6. 肺炎链球菌蛋白质，或其片段或同源物或衍生物，该蛋白质具有 N 末端序列 MELVLPNNYVV (D, A) I (L) D (E) E (Q) EEMMYLDGG (E)，其中括号内的残基代表其前面氨基酸的替换。
7. 核酸分子，其包含或由如下序列构成：
 - (i) 表 1 到 3 所列的任何 DNA 序列或其 RNA 等同物；
 - (ii) 与 (i) 中任何序列互补的序列；
 - (iii) 与 (i) 或 (ii) 中序列编码相同蛋白质或多肽的序列；
 - (iv) 与 (i)、(ii) 和 (iii) 中的任何序列基本相同的序列；
 - (v) 编码表 1 定义的蛋白质的同源物、衍生物或片段的序列。
8. 具有选自表 1 到 3 所示序列的蛋白质或多肽、或其同源物、衍生物和/或片段作为免疫原和/或抗原的用途。
9. 免疫原性和/或抗原性组合物，其包含一或多种选自其序列为表 1 到 3 所示的蛋白质或多肽、或其同源物、或衍生物、和/或任何这些分子的片段。
10. 如权利要求 8 要求的用途，其中抗原和/或免疫原用于疫苗或诊断化验。
11. 如权利要求 10 要求的疫苗，其包含一或多种选自赋形剂、稀释剂、佐剂或类似物质的添加成份。

12. 疫苗组合物，其包含一或多种表 1 到 3 定义的核酸序列。
13. 检测/诊断肺炎链球菌的方法，其包括将要检测的样品与至少一种如表 1 到 3 所定义的蛋白质或多肽、或其同源物、或衍生物、或片段接触的步骤。
14. 能与如表 1 到 3 所定义的蛋白质或多肽、或其同源物、或衍生物、或片段结合的抗体。
15. 如权利要求 14 要求的抗体，其为单克隆抗体。
16. 检测/诊断肺炎链球菌的方法，其包括将要检测的样品与至少一种如权利要求 14 或 15 所定义的抗体接触的步骤。
17. 检测/诊断肺炎链球菌的方法，其包括将要检测的样品与至少一种如权利要求 7 所定义的核酸序列接触的步骤。
18. 确定如表 1 到 3 所定义的蛋白质或多肽是否代表可能的抗菌靶标的方法，其包括失活所述的蛋白质或多肽并确定肺炎链球菌是否仍然存活。
19. 能拮抗、抑制、或以其它方式干扰如表 1 到 3 定义的蛋白质或多肽的功能或表达的物质在制备治疗或预防肺炎链球菌感染的药物中的用途。

分泌型肺炎链球菌蛋白质

本发明涉及源自肺炎链球菌 (*Streptococcus pneumoniae*) 的蛋白质、编码这些蛋白质的核酸分子、上述核酸和/或蛋白质作为抗原/免疫原和在检测/诊断中的用途, 以及筛选蛋白/核酸序列作为可能的抗菌靶标的方法。

肺炎链球菌, 通常称为肺炎球菌, 是一种重要的致病生物。肺炎链球菌感染在发展中国家和发达国家与人类疾病持续重要的关系已有权威性的论述 (Fiber, G.R., *Science*, 265: 1385-1387(1994))。这篇文章指出, 在世界范围内, 该细菌被认为是急性呼吸系统感染最常见的致病菌, 估计每年造成一百万儿童死亡, 多数在发展中国家 (Stansfield, S.K., *Pediatr. Infect. Dis.*, 6: 622 (1987))。在美国, 肺炎球菌仍被认为是细菌性肺炎最常见的致病因素 (Breiman 等人, *Arch. Intern. Med.*, 150: 1401 (1990)), 该病的发病率在幼童、老年人及处于诱病因素状态的患者中尤其高, 诸如无脾、心脏、肺和肾脏疾病患者, 糖尿病、酒精中毒病人或者免疫抑制失调者, 尤其是艾滋病人。这些人群发生肺炎球菌败血症以至脑膜炎的风险很高, 因此死于肺炎球菌感染的风险很大。据信美国每年有超过 50,000 例侵入性肺炎球菌患者 (脑膜炎和菌血症)。肺炎球菌还是中耳炎和鼻窦炎最主要的致病菌, 在发达国家儿童中该菌仍然流行感染, 这导致重大的代价。仅在美国一个国家, 在两岁以下儿童中大概有七百万例肺炎链球菌引起的中耳感染。

由于最近出现了青霉素抗性菌株, 因此更加需要有效预防肺炎球菌感染的措施。据报道, 美国 12 个州的 13 家医院的肺炎球菌分离物中, 有 6.6% 具有青霉素抗性, 其中一些分离株还对其它抗生素具有抗性, 包括第三代环孢菌素 (Schappert, S.M., *Vital and Health Statistics of the Centres for Disease Control/National Centre*

for Health Statistics, 214: 1 (1992))。在一些医院, 青霉素抗性的发生率甚至更高(高达 20%) (Breiman 等人, J. Am. Med. Assoc., 271: 1831 (1994))。在青霉素一直作为肺炎球菌有效的治疗药物几十年后, 由于其抗性菌株在近期的突然出现, 因此这些发现被认为是一种警示。

综上所述, 有必要考虑提高肺炎球菌疾病的预防、控制、诊断或治疗方法。

多种方法用于提供预防肺炎球菌感染的疫苗。例如, 鉴于肺炎球菌有多种血清型(至少 90 种), 因此存在多种困难, 血清型是根据包裹菌体的多糖荚膜的结构划分的。对一种血清型有效的疫苗对其它血清型无效, 这意味着对大多数病例有效的疫苗必须包含全部血清型的多糖抗原。另一个出现的问题是, 当纯化的荚膜多糖(每一种荚膜多糖都决定着血清分型并是主要的保护抗原) 用做疫苗时, 在两岁以下儿童中诱导保护性抗体反应并不可靠, 而这个年龄群患上侵入性肺炎球菌感染和脑膜炎的机率却最高。

为诱导增强的免疫应答, 尤其是 T 细胞依赖性的免疫应答, 一种改进的使用荚膜抗原的方法将多糖与蛋白质缀合。这种方法已经用于针对流感嗜血杆菌 (*Haemophilus influenzae*) 的疫苗的开发。但是有关多种多糖疫苗和以缀合物为基础的疫苗的成本的问题仍然存在。另外, 由于缀合物疫苗提供的血清型覆盖面有限, 因此这种疫苗组合物应当根据地区和人群的不同加以改变。由于施用载体蛋白质总剂量相对较大, 还可能出现载体诱导的免疫抑制或过载的问题。

第三种方法是寻找有潜力成为疫苗候选物的其它抗原性成分。这是本发明的基础。利用一种专门开发的细菌表达系统, 我们能够从肺炎球菌中鉴定出一组蛋白质抗原, 它们与菌体外膜相关或由菌体分泌。

因此, 本发明首先提供了一种肺炎链球菌的蛋白质或多肽, 其具有选自表 1 所示的序列。

本发明的蛋白质或多肽可以基本上纯的形式提供。例如, 它可以基本上不含其它蛋白质的形式提供。

如在此所讨论的，本发明的蛋白质和多肽可用作抗原性物质。这些物质具有“抗原性”和/或“免疫原性”。一般情况下，“抗原性”是指该蛋白质或多肽能用于产生抗体，或者实际上能在受试者体内诱导抗体反应。“免疫原性”是指该蛋白质或多肽能在受试者体内激发保护性的免疫应答。因此，在后一种情况下，该蛋白质或多肽不仅能产生抗体反应，而且，能产生不依赖抗体的免疫应答。

技术人员应当理解，本发明蛋白质或多肽的同源物或衍生物同样可应用于本发明，即用作抗原性/免疫原性物质。例如，包含一个或多个添加、缺失、取代等的蛋白质或多肽包括在本发明的范围内。另外，可以将一个氨基酸用另一个相似“类型”的氨基酸取代。例如用一个疏水性的氨基酸替换另一个疏水性的氨基酸。

技术人员可以使用程序，例如 CLUSTAL™ 程序进行氨基酸序列比较。该程序能比较氨基酸序列并通过在两个序列的合适位置插入空格而找出最佳的比对。对于一个最佳比对，可计算出氨基酸的一致性 or 相似性（一致性加上氨基酸类型的保守性）。象 BLAST_x 这样的程序，能比出最长的相似序列，并给出相似性的分值。因此就有可能得到一组对比序列，它们之间有几个相似的区域，每个相似区域都有不同的分值。本发明考虑了上述两种一致性分析。

对于同源物和衍生物，它们保留的原蛋白质或多肽的抗原性或免疫原性的程度，比它们与此处描述的蛋白质或多肽的一致性程度更重要。但是，提供了合适的同源物或衍生物，其与此处描述的蛋白质或多肽至少要有 60% 的相似性（如前所述）。优选，提供的同源物或衍生物至少有 70% 相似性，更优选，至少有 80% 的相似性。最优选，提供的同源物或衍生物至少有 90% 或者甚至 95% 的相似性。

在另一研究中，同源物或衍生物也可是融合蛋白，融合部分可使纯化更加容易，例如通过有效地给目的蛋白质或多肽加标签。在使用时可能需要移去“标签”，或者融合蛋白本身保留了足够有用的抗原性。

另一方面，本发明还提供蛋白质或多肽，或其同源物或衍生物的

抗原性/免疫原性片段。

对于此处描述的蛋白质或多肽，或其同源物或衍生物的片段，情况略有不同。众所周知，可以筛选抗原性蛋白或多肽来确定抗原表位区，即那些引起蛋白质或多肽抗原性或免疫原性的区域。本领域的技术人员熟知这些筛选方法。所以，本发明中的片段应当包含一个或多个这样的表位区，或者与这些区域有足够高的同源性，以保持抗原性/免疫原性。因此，本发明中的片段可与一致性无关，因为它们可能只与此处描述的蛋白质或多肽，同源物或衍生物的特定部分 100%一致。再次指出，关键是这些片段还保持着抗原性/免疫原性。

对于同源物、衍生物和片段，重要的是它们至少具有一定程度的其所源自的蛋白质或多肽的抗原性/免疫原性。

本发明的蛋白质可利用基因克隆技术以基本上纯化的形式提供。这些技术已经公开，例如在例如在 J. Sambrook 等人，Molecular Cloning 2nd Edition, Cold Spring Harbor Laboratory Press (1989) 中。因此，第二方面，本发明还提供包含或由下述序列构成的核酸分子：

- (i) 表 1 列出的任何 DNA 序列或其 RNA 等同物；
- (ii) 与 (i) 中任何序列互补的序列；
- (iii) 与 (i) 或 (ii) 中序列编码相同蛋白质或多肽的序列；
- (iv) 与 (i)、(ii) 和 (iii) 中的任何序列基本相同的序列；
- (v) 编码表 1 定义的蛋白质的同源物、衍生物或片段的序列。

本发明的核酸分子可能包括多个这样的序列和/或片段。技术人员应能理解，本发明还包括下面举例说明的一些具体的新核酸分子的新的变体。这些变体也包括在本发明的范围内。这些变体可自然发生，例如由于菌种变异。例如，包括添加、替换和/或缺失。另外，特别在使用微生物表达系统时，技术人员可能希望使用特定的用于表达的微生物中公知的偏好密码子设计核酸序列。因此，合成的或非天然发生的变体也包括在本发明的范围内。

上文中使用的术语“RNA 等同物”，是指给定的 RNA 分子的序列与

给定的 DNA 分子序列互补(允许 RNA 中的“U”替换基因密码中的“T”)

当比较核酸序列以确定相似或一致的程度时,技术人员可使用一些程序,例如 BESTFIT 和 GAP(二者均出自 Wisconsin Genetics Computer Group (GCG)的软件包)。以 BESTFIT 为例,它比较两个序列并给出最相似区段的最佳比对。GAP 能比对序列的全长,并通过在两个序列中插入合适的空格而找出最佳的比对。在本发明中,当讨论核酸序列的一致性时,适合通过排列序列的全长进行比较。

优选,基本一致的序列与所述序列至少有 50%的序列一致性,较好的是至少有 75%的序列一致性,更好的是至少有 90%或至少 95%的序列一致性。在某些情况下,序列一致性可达 99%或更高。

术语“基本一致”的合适理解是,该序列与此处描述的任何序列的一致程度高于和现有技术中核酸序列的一致程度。

但应当指出的是,本发明中的核酸序列至少编码某一新基因产物的一部分,编码新基因产物或其新的部分的所有可能的序列均包括在本发明范围内。

核酸分子可以分离或重组的形式存在。它可以连接到载体上,该载体可以整合到宿主中。这些载体和适宜的宿主形成了本发明的另一方面。

因此,例如,使用基于此处提供的核酸序列设计的探针,可鉴定肺炎链球菌的基因。然后它们可用限制性内切酶切出并克隆进载体。该载体可转入合适的宿主进行表达。

通过使用与核酸分子部分序列互补的合适探针,可从肺炎链球菌中获得本发明的核酸分子。利用限制性内切酶或超声波技术,可获得大小合适的片段用作探针。

PCR 技术也可用于扩增目的核酸序列。所以,此处提供的序列数据可用于设计 PCR 中使用的两条引物,这样包括全长基因或其片段的目的序列,可成为目标并被高度扩增。

一般引物的长至少 15-25 个核苷酸。

另外，还可使用化学合成。这可以自动化。相对短的序列可以化学合成，并连接在一起形成长序列。

利用此处描述的细菌表达系统鉴定出了另外一组肺炎链球菌蛋白质。这些是已知的肺炎链球菌蛋白质，但此前没有鉴定为抗原性蛋白质。表 2 列出了这组蛋白质的氨基酸序列，和编码它们的 DNA 序列。这些蛋白质、或其同源物、衍生物和/或片段，也可用作抗原/免疫原。

另一组与 ID-304L1 有一定程度同源性的蛋白质从最近发表的肺炎链球菌基因组中鉴定出来，它们均在 N 端或近 N 端的位置具有下述 23 个氨基酸的高度保守的序列：
MELVLPNNYVV (D, A) I (L) D (E) E (Q) EEMMYLDGG (E)，括号内的残基是前面氨基酸的替换残基。表 3 给出了这些同源物的氨基酸序列和编码它们的 DNA 序列。

所以，另一方面本发明提供肺炎链球菌的蛋白质、或其片段或同源物或衍生物，该蛋白质 N 末端序列为 MELVLPNNYVV (D, A) I (L) D (E) E (Q) EEMMYLDGG (E)。

另一方面，本发明提供了具有选自表 1 到 3 所示序列的蛋白质或多肽、或其同源物、衍生物和/或片段作为免疫原和/或抗原的用途。

另一方面，本发明还提供了一种免疫原性/抗原性组合物，包含一种或多种其序列选自表 1 到 3 所示序列的蛋白质或多肽、或其同源物或衍生物、和/或任何这些分子的片段。在优选的实施方案中，该免疫原性/抗原性组合物是疫苗或用于诊断化验。

作为疫苗时，可包括适当添加的赋形剂、稀释剂、佐剂或类似物质。有很多这些方面的实例为本领域所熟知。

还可使用表 1 和 2 所示的核酸序列制备所谓的 DNA 疫苗。所以，本发明还提供了一种疫苗组合物，包括一种或多种此处所定义的核酸序列。现有技术描述了 DNA 疫苗（例如参见，Donnelly 等人，*Ann. Rev. Immunol.*, 15:617-648 (1997)），技术人员根据本发明，使用此领域描述的技术，可制备和使用 DNA 疫苗。

如此处已讨论的那样，此处描述的蛋白质或多肽、它们的同源物

或衍生物、和/或任何这些分子的片段，可用于肺炎链球菌的检测/诊断方法。这些方法可基于检测抗这些蛋白质的抗体，这些抗体可能在检测对象中出现。因此本发明提供了肺炎链球菌的检测/诊断方法，包括将要检测的样品与至少一种此处描述的蛋白质、或同源物、衍生物或片段接触的步骤。合适的样品是生物样品，如从检测对象中获取的组织样品或血液或唾液样品。

在另一项研究中，此处描述的蛋白质、或其同源物、衍生物和/或片段，可用于产生抗体，抗体反过来可用于检测抗原以及肺炎链球菌。这些抗体组成了本发明的另一面。包括在本发明范围内的抗体可以是单克隆抗体或多克隆抗体。

多克隆抗体可通过以下方式产生，将此处描述的蛋白质、或其同源物、衍生物或片段注射进合适的动物宿主，在动物中（例如小鼠、大鼠、豚鼠、兔、绵羊、山羊或猴），能刺激抗体的产生。如果需要，佐剂可与蛋白质一起使用。众所周知的佐剂包括弗氏佐剂（完全和非完全）和氢氧化铝。然后可利用抗体与此处描述的蛋白质结合的性能纯化它们。

单克隆抗体可由杂交瘤产生。这些可通过将骨髓瘤细胞和产生所需抗体的脾细胞融合，产生永生化的细胞系而形成。因此可使用众所周知的 Kohler 和 Milstein 的技术（Nature 256 (1975)），或此技术的后续改进形式。

制备与特定多肽/蛋白质结合的单克隆和多克隆抗体的技术，目前在本领域中非常成熟。标准免疫学教科书对它们有详细描述，例如：Roitt 等人，*Immunology* second edition (1989)，Churchill Livingstone, London. .

除全抗体外，本发明包括能与此处描述的蛋白质等结合的抗体的衍生物。因此本发明包括抗体片段和合成结构。Dougall 等人给出了抗体片段和合成结构的例子（Tibtec, 12: 372-379 (1994年9月)）。

例如，抗体片段包括：Fab、F(ab')₂和 Fv 片段。Fab 片段在 Roitt 等人[见上文]中有详述。Fv 片段修改后可产生称为单链 Fv (scFv) 分

子的合成结构。这包括共价连接 V_H 和 V_L 区的肽接头，该接头有使分子稳定的作用。其它可用的合成结构包括互补性决定区 (CDR) 肽。这些是包含抗原结合决定簇的合成肽。还可使用肽模拟物 (mimetics)。这些分子一般是构象被限制的有机环，其模拟 CDR 环结构和包括抗原相互作用侧链的结构。

合成结构包括嵌合分子。因此，例如人源化的 (或灵长类化的 (primatised)) 抗体或其衍生物也包括在本发明的范围内。人源化抗体的一个例子是具有人源的抗体骨架区和啮齿类动物的高变区的抗体。制备嵌合抗体的方法有详细论述，例如：Morrison 等人，PNAS, 81: 6851-6855 (1984) 和 Takeda 等人，Nature, 314: 452-454 (1985)。

合成结构还包括带添加部分的分子，添加部分使分子具有除抗原结合性质外的其他所需性质。例如该部分可能是一种标记 (例如荧光或放射性标记)。另外，它可能是一种药物活性成分。

抗体或其衍生物，可用于肺炎链球菌的检测/诊断。因此，另一方面本发明提供一种肺炎链球菌的检测/诊断方法，它包括将要检测的样品与抗体接触的步骤，该抗体能结合此处描述的一种或多种蛋白质、或其同源物、衍生物和/或片段。

另外，可使用所谓的“亲和体” (Affibodies)。这些是选自 α -螺旋细菌受体结构域组合文库 (Nord 等人 Nature Biotechnology, 15: 772-7 (1997)) 的结合蛋白质。因此，可通过组合的方法选择能与不同靶蛋白质特异结合的小蛋白质结构域。

还将明确的是，此处描述的核酸序列可用于检测/诊断肺炎链球菌。因此，另一方面，本发明还提供了一种检测/诊断肺炎链球菌的方法，该方法包括将要检测的样品与至少一种此处描述的核酸序列接触的步骤。合适的样品是生物样品，如从被检测对象中获取的组织样品或血液或唾液样品。这些样品在用于本发明的方法之前可进行预处理。例如，可处理样品以提取 DNA。然后，以此处描述的核酸序列为基础的 DNA 探针 (即，一般是这些序列的片段)，可用于检测肺炎链球菌的核酸。

另外，本发明还提供：

(a) 针对肺炎链球菌免疫接种受试者的方法，包括对受试者施用本发明的蛋白质或多肽、或其衍生物、同源物或片段，或本发明的免疫组合物的步骤；

(b) 针对肺炎链球菌免疫接种受试者的方法，包括对受试者施用如此处所定义的核酸分子的步骤；

(c) 预防或治疗肺炎链球菌感染的方法，包括对受试者施用本发明的蛋白质或多肽、或其衍生物、同源物或片段，或本发明免疫组合物的步骤；

(d) 预防或治疗肺炎链球菌感染的方法，包括对受试者施用如此处所定义的核酸分子的步骤；

(e) 用于检测/诊断肺炎链球菌感染的试剂盒，其包括本发明的一或多种蛋白质或多肽、或其同源物、衍生物或片段，或本发明的抗原性组合物；和

(f) 用于检测/诊断肺炎链球菌感染的试剂盒，其包括一或多种如此处所定义的核酸分子。

考虑到我们已经鉴定出了一组重要的蛋白质，这些蛋白质是抗菌治疗的可能靶标。但是，必需确定每一种蛋白质对细菌的存活是否重要。因此，本发明还提供了一种确定如此处描述的蛋白质或多肽是否代表可能的抗菌靶标的方法，该方法包括拮抗、抑制、或以其它方式干扰该蛋白质的功能或表达并确定肺炎链球菌是否仍然存活。

使蛋白质失活的合适方法是实施选择性基因敲除，即，阻止该蛋白质的表达并确定其结果是否为致死性的变化。进行基因敲除的合适方法在下列论文中有描述：Li 等人，P. N. A. S., 94:13251-13256 (1997) 和 Kolkman 等人，J. Bacteriol., 178:3736-3741 (1996)。

最后一个方面，本发明提供了一种物质在制备治疗或预防肺炎链球菌感染的药物中的应用，其中该物质能够拮抗、抑制、或以其它方式干扰本发明蛋白质或多肽的功能或表达。

如前所述，我们使用细菌表达系统来鉴定蛋白质，其中那些蛋白质是表面相关的、分泌的或输出的并因此用作抗原或抗菌靶标。

蛋白质分泌/输出的必需信息在细菌中已经被深入研究过。多数情况下，蛋白质输出需要一个存在于前体蛋白质 N 末端的信号肽，通过信号肽将蛋白质引入细胞质膜上的转运机器。在转运过程中或转运之后，信号肽被膜相关的信号肽酶去除。最终，蛋白质的定位（即，是否被分泌、完整的膜蛋白或附着于细胞壁）由序列决定，而不是前导肽本身。

我们对定位于表面或输出的蛋白质特别感兴趣，因为它们很可能作为抗原用于疫苗、作为诊断试剂或作为新化学物质的治疗靶标。因此我们在乳酸乳球菌 (*Lactococcus lactis*) 中开发了一个筛选载体系统，可分离和鉴定编码输出蛋白质的基因。下面我们提供一个代表性的例子，说明如何鉴别和表征给出的肺炎链球菌新的表面相关蛋白质。筛选载体结合了缺少自身输出信号的葡萄球菌核酸酶基因 *nuc* 作为分泌报告分子。葡萄球菌核酸酶是自然分泌的、热稳定的单体酶，它在一系列革兰氏阳性菌中有效表达和分泌 (Shortle, *Gene*, 22:181-189 (1983); Kovacevic 等人, *J. Bacteriol.*, 162:521-528 (1985); Miller 等人, *J. Bacteriol.*, 169:3508-3514 (1987); Liebl 等人, *J. Bacteriol.*, 174:1854-1861 (1992); Le Lori 等人, *J. Bacteriol.*, 176:5135-5139 (1994); Poquet 等人, *J. Bacteriol.*, 180:1904-1912 (1998))。

最近 Poquet 等人 (1998, 见上文) 描述了一种筛选载体，该载体结合了缺少自身信号前导 (signal leader) 的 *nuc* 基因作为报告分子，用于鉴定革兰氏阳性菌的输出蛋白质，并在乳酸乳球菌中应用。该载体 (pFUN) 除带有在大肠杆菌 (*Escherichia coli*) 和其它革兰氏阴性菌中促进复制的 ColE1 复制子外，还带有在大范围革兰氏阳性菌中发挥功能的 pAMB1 复制子。该载体上独特的克隆位点可用于产生转录和翻译的融合体，该融合体在克隆的基因组 DNA 片段和缺失自身分泌信号前导的截断的 *nuc* 基因之间形成。*nuc* 基因是理想的报告基

因，因为核酸酶的分泌容易被简单和灵敏的平皿实验检测；分泌核酸酶的重组克隆形成粉红色的晕圈，而对照克隆仍保持白色（Shortle, (1983), 见上文；Le Lori 等人, (1994), 见上文）。

因此参考下面代表性的例子来描述本发明，这些例子提供了如何把此处描述的蛋白质、多肽和核酸序列鉴定为抗原性靶标的细节。

这里我们描述了三种报告载体的构建，和它们在乳酸乳球菌中鉴定和分离肺炎链球菌中编码分泌或表面相关蛋白质的基因组 DNA 片段的用途。另外，使用 DNA 印迹杂交实验证明在一系列肺炎链球菌株中存在疫苗候选基因。现在要参考实施例来描述本发明，实施例在任何情况下都不作为限制本发明的解释。

实施例 1

(i) pTREP1-nuc 系列报告载体的构建

(a) 表达质粒 pTREP1 的构建

pTREP1 质粒是高拷贝数（40-80 每细胞） θ -复制革兰氏阳性质粒，它衍生自 pTREX 质粒，pTREX 本身衍生自以前公开的 pIL253 质粒。pIL253 带有 pAM β 1 复制子，该复制子有广泛的革兰氏阳性宿主范围（Simon 和 Chopin, *Biochimie*, 70:559-567 (1988)），并且乳酸乳球菌的性因子（sex-factor）不能转移该质粒。pIL253 还缺少 *tra* 功能，该功能为通过结合母体质粒（conjugative parent plasmids），例如 pIL501 转移或有效移动所必需的。肠球菌复制子 pAM β 1 以前已被转移到多种细菌中，包括链球菌、乳酸杆菌（*Lactobacillus*）和芽孢杆菌（*Bacillus*）以及丙酮丁醇梭状芽孢杆菌（*Clostridium acetobutylicum*）（Oultram 和 Klaenhammer, *FEMS Microbiological Letters*, 27:129-134 (1985)；Gibson 等人, (1979)；LeBlanc 等人, *Proceedings of the National Academy of Science USA*, 75:3484-3487 (1978)），这说明 pAM β 1 可应用于广泛的宿主范围 pTRET1 质粒代表组成型转录载体。

pTREX 载体的构建如下。通过将两条互补的寡核苷酸退火并用 Tfl

DNA 聚合酶延伸生成一条人造 DNA 片段, 该片段包括推定的 RNA 稳定序列、翻译起始区 (TIR)、插入目的基因的多克隆位点和转录终止子。为便于克隆, 有义和反义寡核苷酸分别在其 5' 末端带有 NheI 和 BamHI 的识别位点。此片段被克隆在 pUC19NT7 的 XbaI 和 BamHI 位点之间, pUC19NT7 衍生自 pUC19, 后带有来自 pLET1 的 T7 表达盒 (Wells 等人, *J. Appl. Bacteriol.*, 74:629-636 (1993)), 该表达盒克隆在 EcoRI 和 HindIII 位点之间。构建结果称为 pUCLEX。然后先用 HindIII 酶切, 平端化后再用 EcoRI 酶切以移出 pUCLEX 完整的表达盒, 再克隆到 pIL253 的 EcoRI 和 SacI (平端化的) 位点之间, 从而形成 pTREX 载体 (Wells 和 Schofield, *In Current advances in metabolism, genetics and applications-NATO ASI Series*, H 98:37-62 (1996))。推定的 RNA 稳定序列和 TIR 来自大肠杆菌 T7 噬菌体序列, 并对其中一个核苷酸位点进行了修饰以增强 Shine Dalgarno (SD) 基序与乳酸乳球菌核糖体 16S RNA 的互补能力 (Schofield 等人, pers. coms. University of Cambridge Dept. Pathology.)。

后来被称为 T7 的具有启动子活性的乳酸乳球菌 MG1363 染色体 DNA 片段, 被克隆进表达盒的 EcoRI 和 BglIII 位点之间, 从而产生 pTREX7。以前已利用启动子探针载体 pSB292 分离到这个有效的启动子区域 (Waterfield 等人, *Gene*, 165:9-15 (1995))。根据制造商的说明使用 Vent DNA 聚合酶 PCR 扩增该启动子片段。

然后 pTREP1 载体的构建如下。通过将两条部分重叠的互补的合成寡核苷酸一起退火, 并根据制造商的说明用测序酶延伸, 生成一条人造 DNA 片段, 该片段包括转录终止子、pUC 上游测序引物、启动子多克隆位点区和通用翻译终止序列。为便于克隆到 pTREX7 中, 有义和反义 (pTREP_F 和 pTREP_R) 寡核苷酸分别在其 5' 末端带有 EcoRV 和 BamHI 的识别位点。转录终止子来自芽孢杆菌青霉素酶基因, 已被证明在乳球菌中有效 (Jos 等人, *Applied and Environmental Microbiology*, 50:540-542 (1985))。转录终止子被认为是必需的, 因为观察到 pTREX 载体在表达目的基因时渗漏, 并认为是起始区段隐藏的启动子活性造

成的结果 (Schofield 等人, pers. coms. University of Cambridge Dept. Pathology.)。包含的 pUC 上游测序引物能引导克隆 DNA 片段的测序。包含的翻译终止序列以 3 种不同的阅读框架编码终止密码子, 以防止载体基因和克隆的 DNA 片段之间出现翻译融合体。pTREX7 载体先用 EcoRI 消化, 并根据制造商的说明用 T4 DNA 聚合酶 (NEB) 的 5'-3' 聚合酶活性平端化。EcoRI 消化并平端化的 pTREX7 载体再用 BglII 消化以移出 P7 启动子。然后, 由合成寡核苷酸退火获得的人造 DNA 片段用 EcoRV 和 BamHI 消化, 并克隆进 EcoRI (平端化) - BglII 消化的 pTREX7 载体, 形成 pTREP。再将称为 P1 的乳酸乳球菌 MG1363 染色体启动子克隆到 pTREP 表达盒的 EcoRI 和 BglII 位点之间, 形成 pTREP1。这个启动子也被 Waterfield 等人 ((1995), 见上文) 利用启动子探针载体 pSB292 分离和表征。最初根据制造商的说明使用 Vent DNA 聚合酶 PCR 扩增 P1 启动子片段, 并克隆进 pTREX 载体作为 EcoRI - BglII DNA 片段。通过限制性内切酶消化, 从 pTREP1 中移出带有 P1 启动子的 EcoRI - BglII 片段, 并克隆进 pTREP 中 (Schofield 等人, pers. coms. University of Cambridge Dept. Pathology.)。

(b) PCR 扩增金黄色葡萄球菌 (*S. aureus*) *nuc* 基因

金黄色葡萄球菌 *nuc* 基因的核苷酸序列 (EMBL 数据库登录号 V01281), 用于设计进行 PCR 扩增的合成寡核苷酸引物。设计的引物用于扩增称为 nucA 的 *nuc* 基因的成熟形式, 其由称为 Snase B 的分泌型前体肽被蛋白水解切除 N 末端 19 到 21 个氨基酸后生成 (Shortle, (1983), 见上文)。设计了三条有义引物 (nucS1、nucS2 和 nucS3, 附录 1), 每一条引物以相对 *nuc* 基因不同阅读框具有 EcoRV 或 SmaI 的平末端限制性内切酶切割位点。另外, 有义和反义引物的 5' 末端分别带有 BglII 和 BamHI 以方便克隆到由 BamHI 和 BglII 切割的 pTREP1。附录 1 中给出了所有引物的序列。使用上述各有义引物结合反义引物 PCR 扩增三个编码核酸酶基因成熟形式 (NucA) 的 *Nuc* 基因 DNA 片段。使用金黄色葡萄球菌基因组 DNA 模板、Vent DNA 聚合酶 (NEB) 和制造商推荐的条件 PCR 扩增 *Nuc* 基因片段。起始步骤为 93°C 变性 2 分钟,

然后 93℃ 变性 45 秒、50℃ 退火 45 秒、和 73℃ 延伸 1 分钟，进行 30 个循环，最后步骤为 73℃ 延伸 5 分钟。用 Wizard clean up column (Promega) 纯化 PCR 扩增产物以去除未结合的核苷酸和引物。

(a) pTREP1-nuc 载体的构建

如 (b) 部分所述纯化的 nuc 基因片段使用标准条件用 BglIII 和 BamHI 消化，并与 BamHI 和 BglIII 切割的及脱磷酸的 pTREP1 连接，形成 pTREP1-nuc1、pTREP1-nuc2 和 pTREP1-nuc3 系列的报告载体。使用制造商提供的试剂和缓冲液或使用标准条件 (Sambrook 和 Maniatis, (1989), 见上文) 进行普通分子生物学技术操作。在各 pTREP1-nuc 载体中表达盒包括转录终止子、乳球菌启动子 P1、独特的克隆位点 (BglIII、EcoRV 或 SmaI) 和后续的 nuc 基因的成熟形式及第二个转录终止子。注意 nuc 基因翻译和分泌必需的序列被有意排除在这一结构之外。这些成分只能由合适消化的外源 DNA 片段 (代表靶细菌的) 提供，该片段可被克隆到在 nuc 基因上游紧邻的独特的限制性位点。

在拥有启动子方面，pTREP1-nuc 载体与 Poquet 等人 (1998), 见上文) 描述的 pFUN 载体不同，其用于直接在乳酸乳球菌中直接筛选 Nuc 的活性来鉴定乳酸乳球菌的输出蛋白。由于 pFUN 载体在 nuc 开放阅读框架上游没有启动子，因此克隆的基因组 DNA 片段除带有 Nuc 翻译起始和分泌所需的成分外，还必需提供转录信号。这一限制可能会阻止远离启动子的基因的分离，例如位于多顺反子操纵子中的基因。而且不能保证来自其它种类细菌的启动子在乳酸乳球菌中会被识别和发挥功能。某些启动子在天然宿主内可受到严紧调控，但在乳酸乳球菌中却不受这种调控。相反，带有 P1 启动子的 pTREP1-nuc 系列载体能保证无启动子的 DNA 片段 (或 DNA 片段含在乳酸乳球菌中无活性的启动子序列) 仍被转录。

(ii) 肺炎链球菌分泌蛋白质的筛选

从肺炎链球菌中分离的基因组 DNA 用限制性内切酶 Tru9I 消化。之所以使用这个识别序列为 5'-TTAA-3' 的酶，是因为它能有效切割富含 A/T 的基因组并随机产生在优选尺寸范围 (一般平均为 0.5 - 1.0

kb) 的基因组 DNA 片段。这个尺寸范围是优选的, 因为使用 P1 启动子转录新基因序列的可能性有提高。但是, 并非所有情况下都需要 P1 启动子, 因为许多链球菌的启动子可能在乳酸乳球菌中被识别。从 Tru9I 部分消化的肺炎链球菌基因组 DNA 中纯化不同大小范围的 DNA 片段。由于 Tru9I 限制性内切酶产生交错末端, 所以 DNA 片段在连接到 EcoRV 或 SmaI 切割的 pTREP1-nuc 载体之前必需变为平末端。通过使用 Klenow 酶的 5' - 3' 聚合酶活性进行部分填充的酶反应实现这一点。简言之, Tru9I 消化的 DNA 溶解在溶液中(一般总体积在 10-20 μ l 之间), 其补充有 T4 DNA 连接酶缓冲液 (New England Biolabs; NEB) (1 \times) 和 33 μ M 每一种需要的 dNTPs, 此处为 dATP 和 dTTP。加入 Klenow 酶 (1 单位 Klenow 酶 (NEB) 每 μ g DNA), 反应在 25 $^{\circ}$ C 孵育 15 分钟。75 $^{\circ}$ C 孵育该混合物 20 分钟以终止反应。然后加入 EcoRV 或 SmaI 消化的 pTREP1-nuc 质粒 DNA (一般在 200-400 ng 之间)。然后在混合物中补充 400 单位的 T4 DNA 连接酶 (NEB) 和 T4 DNA 连接酶缓冲液 (1 \times), 16 $^{\circ}$ C 孵育过夜。连接混合物直接用 100% 的乙醇和 1/10 体积的 3M 醋酸钠 (pH 5.2) 沉淀, 用于转化乳酸乳球菌 MG1363 (Gasson, 1983)。或者, pTREP-nuc 载体基因克隆位点中还带有用于克隆的 BglIII 位点, 例如用于克隆 Sau3AI 消化的基因组 DNA 片段。

乳酸乳球菌转化克隆在脑心浸液琼脂上生长, 通过覆盖一层甲苯胺蓝-DNA-琼脂 (0.05 M Tris pH 9.0, 每升 10 克琼脂, 每升 10 克 NaCl, 0.1 mM CaCl₂, 0.03% wt/vol 鲑鱼精 DNA 和 90 mg 甲苯胺蓝 O 染料), 可以检测出核酸酶分泌 (Nuc⁺) 克隆, 基本上如 Shortle 等人 (Shortle, 1983, 见上文和 Le Loir 等人, 1994, 见上文) 所描述。然后平板在 37 $^{\circ}$ C 孵育 2 小时。核酸酶分泌克隆形成易于鉴别的粉红色晕圈。从 Nuc⁺重组乳酸乳球菌克隆中分离质粒 DNA, 插入 DNA 用附录 1 中描述的 NucSeq 测序引物在一条链上测序, 这直接测序穿过插入的 DNA 测序。

(iii) 从肺炎链球菌中分离编码输出型蛋白质的基因

使用核酸酶筛选系统已鉴定出大量推定的编码肺炎链球菌输出型

蛋白质的基因。现在进一步鉴定它们以去除人为现象。使用几个参数分析用该筛选系统鉴定的序列。

1. 用软件程序 Sequencher (Gene Codes Corporation) 和 DNA Strider (Marck, *Nucleic Acids Res.*, 16:1829-1836 (1988)), 分析所有推定的表面蛋白质的前导/信号肽序列。细菌信号肽序列有共同的特点。它们的特点是, 在短的带正电的 N 末端 (N 区) 之前, 紧随着一段疏水残基 (中段-h 区), 然后是带有切割位点的极性更强的 C 末端区 (c-区)。利用计算机软件可给出推定的蛋白质的亲水性质, 并易于鉴定出前导肽序列典型的非常明显的疏水区 (h-区)。另外, 检查序列存在或缺失可能的核糖体结合位点 (Shine Dalgarno 基序), 该位点为推定的 nuc 报告融合蛋白质翻译起始所必需。

2. 使用公开提供的数据库 (OWL-proteins, 包括 Swissprot 和 GenBank translations), 将所有推定的表面蛋白质序列与全部蛋白质/DNA 序列比较。这使我们能鉴定出与已知基因相似的序列或已知某些功能的基因的同源物。因此可预测某些利用 LEEP 系统已鉴定出的基因的功能, 而且可无疑地确定该系统可用于鉴定和分离表面相关蛋白质的基因序列。所以我们还应该能够证实这些蛋白质确实是表面相关的, 而不是人为现象。LEEP 系统已用于鉴定疫苗和治疗用的新基因靶标。

3. 一些基因鉴定的蛋白质 (Some of the genes identified proteins) 没有典型的前导肽序列, 也与数据库中的任何 DNA/蛋白质序列无同源性。实际上这些蛋白质可能表明了我们的筛选方法的重要优点, 即, 分离非典型的表面相关蛋白质, 其可能被前述的基于序列同源性搜索的所有筛选方案或方法错过。

所有实例中, 开始仅获得部分基因序列。表 2 列出的全长基因均是通过参考 TIGR 肺炎链球菌数据库 ([www@tigr.org](http://www.tigr.org)) 获得的。这样, 通过将最初获得的部分序列与数据库比较, 我们能鉴定出全长基因序列。因此, 如此处所述, 两组基因被清楚地鉴定, 即, 一组基因编码以前未鉴定的肺炎链球菌蛋白质 (表 1), 第二组编码已知的肺炎链

球菌蛋白质，但不知道这些蛋白质可作为抗原（表2）。

最近又有两个肺炎链球菌基因组被测序并发表了相关信息，随后这些信息由NCBI数据库提供。“临床分离的19F型肺炎链球菌基因组序列注释草图”由Dopazo等人于2001年7月发表(*Microbial Drug Resistance*, 卷7, pp99-125)。“R6株肺炎链球菌基因组”由Hoskins等人于2001年10月发表(*Journal of Bacteriology*, 卷183, pp5709-5717)。通过BLAST分析，从这些基因组中鉴定出了ID-304L1的同源物，它们都在N末端或近N端有一个高度保守的23个氨基酸的序列：MELVLPNNYVV(D, A)I(L)D(E)E(Q)EEMMYLDGG(E)，此处括号内的残基是前面氨基酸的替换。表3给出了这些同源物的序列。

实施例2

肺炎链球菌不同分离物中ID-304L变异体的保守性和变异性

分析了血清型3的肺炎链球菌株ATCC49619中基因ID304L1和ID305的存在情况。根据表1列出的已知核酸序列设计寡核苷酸引物，PCR扩增这些目的基因。

(i) 扩增和标记特异目的基因作为DNA印迹分析的DNA探针

设计寡核苷酸引物用于扩增相应的基因特异的DNA探针(附录2)。根据制造商的说明，用*Pfu Turbo* DNA聚合酶(stratagene)PCR扩增特异的基因靶标(ID304L1和ID305)。典型反应在50 μl体积中进行，包括DNA模板100 ng、十分之一体积的酶反应缓冲液、引物各100 ng、dNTP各200 μM和1.25单位的*Pfu Turbo* DNA聚合酶。典型反应包括：95℃起始变性3分钟，然后是94℃60秒的单循环，然后50℃60秒进行30个循环。然后是72℃2分钟的单循环，接着是72℃最后延伸10分钟。

用QIAquick PCR Purification Kit (Qiagen)纯化所有PCR扩增产物。然后确定ATCC49619株中ID304L1和ID305同源物的存在。

为用作DNA探针，纯化从ID304L1中扩增的基因DNA片段，然后根据制造商的说明，用DIG Nucleic Acid High Prime Labelling Kit

(Roche) 对其标记地高辛。

(ii) DNA 印迹杂交分析 B 组链球菌基因组 DNA

用 DNA 印迹分析确定由 LEEP 系统分离的新肺炎链球菌基因的交差血清型保守性。图 1 给出了分析中使用的肺炎链球菌株。

从肺炎链球菌中分离基因组 DNA，对其进行 ID304L1 衍生的基因靶标的保守性分析。根据制造商的说明，用 HindIII 限制性内切酶 (Roche) 消化合适浓度的 DNA，并用琼脂糖凝胶电泳分析。DNA 样品琼脂糖凝胶电泳后，凝胶在 0.5M NaOH - 1.5M NaCl 中变性 20 分钟、在 0.5M Tris.HCl (pH 7.5) - 1.5M NaCl 里中和 40 分钟，然后经过夜毛细印迹将 DNA 转移到 Hybond N+膜上 (Amersham)。该方法基本上如“用于滤膜杂交的地高辛系统用户使用指南” (Boehringer Mannheim, 1995) 所描述，在下面贮有 20×SSC (柠檬酸钠盐) 的平台上，用 Whatman 3MM (滤纸) 为芯。滤膜在 2×SSC 中短暂漂洗并保存于 -20℃。

当使用 DIG Nucleic Acid Detection Kit 时，滤膜用地高辛标记的 DNA 探针预杂交和杂交，并用 Boehringer Mannheim 推荐的条件洗涤。滤膜在地高辛 “EasyHyb” 中于 42℃ 预杂交一小时。在加入杂交缓冲液 (地高辛 “EasyHyb”) 之前，地高辛标记的 DNA 探针在 100℃ 变性 10 分钟，并于冰上冷却。在 Hybaid 小杂交炉的旋转 Hybaid 管中，进行杂交过夜。用 2×SSC-0.1% SDS 室温 5 分钟的条件洗涤滤膜两次，可去掉未结合的探针。为提高严谨性，滤膜用 0.5×SSC-0.1% SDS 68℃ 15 分钟的条件洗涤两次。使用该 DIG Nucleic Acid Detection Kit (Roche) 免疫检测特异结合的地高辛标记 DNA 探针。

DNA 印迹杂交显示，在多数分析的菌株中存在 ID-304L1 同源物。泳道 12，探针从其中扩增，只有一条很弱的条带，但这归因于凝胶中只加入了低水平的 DNA。这也可用于解释泳道 5 没有条带的情况，该泳道的背景显著低于其它泳道。有些菌株出现两个条带，这说明可能有一个以上的同源物存在 (如在 G54 和 R6 株中发现的那样)。多数临床相关菌株中存在此蛋白质的基因，说明它是一个保守性的蛋白质，因此是一个好的疫苗候选物。

附录 I - LEEP 筛选的寡核苷酸引物

nucS1

Bgl II Eco RV

5'- cgagatctgatatctcacaaacagataacggcgtaaatag -3'

nucS2

Bgl II Sma I

5'- gaagatcttccccgggatcacaaacagataacggcgtaaatag -3'

nucS3

Bgl II Eco RV

5'- cgagatctgatatccatcacaaacagataacggcgtaaatag -3'

nucR

Bam HI

5'- cgggatccttatggacctgaatcagcgttgtc -3'

NucSeq

5'- ggatgctttgtttcaggtgtatc -3'

pTREPf

5'- catgatatcgggtacctcaagctcatatcattgtccggcaatgggtgggctttttgttttagcggataa
caatttcacac -3'

pTREP r

5'- gcggatccccgggcttaattaatgttaaacactagtcgaagatctcgcgaattctcctgtgtgaaatt
gttatccgcta -3'

pUCF

5'- cgccagggttttccagtcacgac -3'

VR

5'- tcaggggggcggagcctatg -3'

V1

5'- tcgtatgtgtgtggaattgtg -3'

V2

5'- tccggctcgtatgtgtggaattg -3'

附录 2 PCR 分析和 DNA 印迹的寡核苷酸引物

设计引物以提供后续克隆中使用的限制性内切酶位点（下面以粗体形式给出）。下划线 CGC 夹（clamps）可使限制性内切酶的结合能力更大。

ID 305 Bam5' GGC **GGATCC** ATA AAC GAA GAA ATA AGC AAG GAA GC

ID 305 Hind3' GGC **AAGCTT** TTA GAT TTC TCT GGT CAT ATC

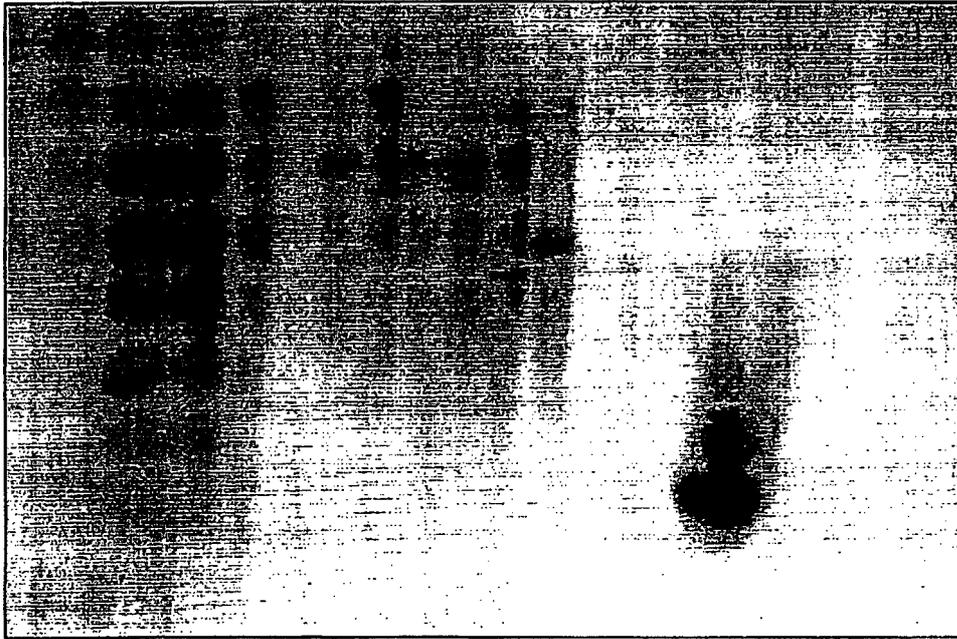
ID 304 L1 Bam5' GGC **GGATCC** AAA CAA TTT CAA CTA AGG AGG

LID 304 L1 Hind3' GGC **AAGCTT** TCA TCT TAC TGT CGC AGA TAT G

ID304L1基因在一系列血清型中的保守性

用HindIII (Roche)完全消化来自各菌株的基因组DNA,并在1.0%琼脂糖(凝胶)上于12伏电泳20小时,通过DNA印迹转移到Hybond N+(Amersham)膜上,与地高辛标记的LID-304基因探针杂交。使用DIG Nucleic Acid Detection Kit (Boehringer Mannheim)鉴定特异结合的DNA探针。

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16



- 1 肺炎链球菌血清型5的临床分离物
- 2 肺炎链球菌血清型18C的临床分离物
- 3 肺炎链球菌血清型23F的临床分离物
- 4 肺炎链球菌血清型7F的临床分离物
- 5 肺炎链球菌血清型1的临床分离物
- 6 肺炎链球菌血清型6B的临床分离物
- 7 肺炎链球菌血清型4的临床分离物
- 8 肺炎链球菌血清型3的临床分离物
- 9 肺炎链球菌血清型19F的临床分离物
- 10 肺炎链球菌血清型9V的临床分离物
- 11 肺炎链球菌血清型14的临床分离物
- 12 肺炎链球菌株ATCC 49619 (血清型3)
- 13 粘膜炎莫拉菌 (*Moraxella catarrhails*) DNA
- 14 地高辛标记的 λ HindIII标志 (marker)
- 15 来自ATCC 49615的LID304L1基因

图1

表1

| | |
|-----------|--|
| ID-303A | MAGNSFHLTLTSVVSQAGQQLRHNSPI |
| ID-303B | ATGGCAGGCAATTCCTTTACCTAACTCTCACTTCTGTATCTCAGGCAGGA CAACAAACGCTTCGACACAATCACAGTCTATT |
| ID-305A | MINEEISKEAGQAAQTIIISYTIKATKESINLEKEIRKKMNETLEKANGNLK SLMGDEMNIKLDLYKKGQLENI SIDQIDLKDLKKELNKLGVSFSVMKNKESK NYEIFFQAKDIKVMYAFKQVIAKENKKEKESILKQIKKYKDLKSKNKDKTK EKGKRKVKPNKKDMTREI |
| ID-305B | ATGATAAACGAAGAAATAAGCAAGGAAGCAGGTCAAGCAGCACAAACCATA ATATCATAACAATAAAGGCAACAAAAGAATCAATCAATTTAGAAAAAGAA ATAAGAAAAAGATGAATGAACTTTAGAAAAAGCAAAATGGAACTTAAAA AGTCTTATGGGCGATGAAATGAAAATAAAAGACCTCTACAAGAAAGGACAA CTAGAAAATATAAGCATAGATCAAATCGACCTCAAAGACTTAAAAAAGAA CTAAACAAACTTGAGTAAGTTTCTCAGTAATGAAAAACAAAGAAAGCAAA AACTATGAAATATCTTCCAGCCAAAGACATAAAAGTAATGGAATATGCC TTTAAGCAAGTCATAGCCAAGGAAAATAAAAAAGAAAAAGAAAGTATCCTA AAACAAATAAAGAAATACAAAGACCTATCCAAAAACAAAGATAAGACAAAA GAAAAAGGAAAAGGAAAGTAAAAGAAAAAGTAAACCAAACAAAAAAGAT ATGACCAGAGAAATC |
| ID-306A | MKVSKKITLFLSLSFAGFVLLTLPQAGKAFELKEDWAFKGGIRYENGKVS KINNGYEVNIKVLDPSTSAIEWTVRLNGEKQNTNFLAEERTVSKTEDKGR FLHFYIPYGYRGDIVVEAKSGNEVKTWSTKVVDDVYSDSAKSGYFILDGE QILESSWDSVNESYIATLPTVTSKGTVVAVREKGTNLNI |
| ID-306B | ATGAAAGTATCAAAAAAATTACACTATTTAGTTTGTCTTTTGCAGGTTT TGTTTTATTGACTTTACCTCAAGCAGGAAAGGCTTTGAACTTAAAGAAG ACTGGGCATTTAAAGGTGGCATTTCGATACGAGAATGGGAAAGTCAGCAA ATTAATAATGGATATGAAGTAAATATTAAGTGTAGATTTACCTAGTAC TAGCGCAATCGAATGGACAGTTAGATTGAATGGAGAAAAGCAAAATACTA ACTTCTTAGCGGAGGAAAGAACTGTATCTAAAACCTGAAGATAAGGGACGT TTCTTGCACTTTTATATCCCCTATGGATATCGTGGGGATATTGTAGTAGA GGCTAAGAGTGGAAACGAAGTGAAGACTTGGTCTACTAAGGTAGTTGACG ATGTTTATTAGATTCTGCTAAGAGTGGCTACTTTATCTCGATGGGGAA CAAATCTTAGAAAGTTCATGGGATTCGGTAAATGAGTCTTATATGCAAC GCTTCCAACGTAAACATCAGGAAAAACTGTTGTTGCTTGGCGTAAAAAG GAACTCTTAATTTAATT |
| ID-304L1A | MELVLPNNYVVLEQEEMMYLDGGFSIPRWPVATAINIAFNGLGGGAI SLVRNYIRNYGLRRVTSIAGAAARYVGVVRVANRVAGFALSAINGF AAWMSIGDAITTIWANNDVNRDPNLNLW |
| ID-304L1B | ATGGAACCTCGTATTACCAAATAATTATGTTGTTCTTGAGCAAGAAGAGAT GATGTATCTTGATGGGGATTTCTATTCCGAGATGGCCTGTTGCAACAG CCATTAATATAGCTTTTAATGGTGTGTTTAGGTGGAGGACAAATCAGTCTA GTTAGAAATTATATTCGTAATTATGGTTTGC GGCGAGTTACAAGCGCAAT TGCTGGAGCAGCTGCAAGATATGTTGGGTACGAGTTGCAAATAGAGTGG CAGGATTTGCACCTGCTGCTATTAATGGATTTGCAGCTTGATGTCAATT GGCGATGCTATTACAACAATCTGGCCAACAATGATGTAAATAGGAGAGA CCCAAATTTAAACGCCTTGTTGTTAA |

表2

括号内的残基代表其前一个残基的替换。
扩增出IUPAC核酸多义密码子

ID-204A MKDTFKNVLSFEFWQKFGKALMVVIAVMPAAGLMISIGKSIVMINPTFAP
LVITGGILEQIGWGVIGNLHILFALAIGGSWAKERAGGAFAAGLAFILIN
RITGTIFGVSGDMLKNPDAMVTTFFGGSIKVADYFISVLEAPALNMGVVF
GIISGFVGATAYNKYYNFRKLPDALSFFNGKRFVPFVVILRSAIAAILLA
AFWPVVQTGINNFGIWIANSQETAPILAPFLYGTLERLLLPFGLHHMLTI
PMNYTALGGTYDILTGAAKGTQVFGQDPLWLAWVTDLVNLKGT DASQYQH
LLDTVHPARFKVQMGISFGILMGVIVAIYRNV DADKKHKYKGMMIATAL
ATFLTGVTEPIEYMFNFIAATPMYLVYSLVQGA AFAMADVNLRMHSFGSI
EFLTRTPIAISAGIGMDIVNFVWVTVLFAVIMYFIANFMIQKFNYATPGR
NGNYETAEGSEETSSEVKVAAGSQAVNIINLLGGRVNI VDVDACMTRLRV
TVKDADKVGNAEQWKAEGAMGLVMKGGVQAIYGP KADILKSDIQDILDS
GEIIPETLPSQMTEAQONTVHFKDLTEEVYSVADGQVVALEQVKDPVFAQ
KMMGDGFAVEPANGNIVSPVSGTVSSI FPTKHAFGI VTEAGLEVLVHIGL
DTVSLEGKPFVTVHVAEGQKVAAGDLLVTADLDAIRAAGRETSTVVVFTNG
DAIKSVKLEKTGSLAAKTAVAKVEL*

ID-204B GGTAAGGCTTTGATGGTAGTTATCGCGGTTATGCCGGCTGCTGGTTGATG
ATTTCAATCGGTAAGTCTATCGTGATGATTAACCAACCTTTGCACCACTT
GTCATCACAGGTGGAATCTTGAGCAAATCGGTTGGGGGTTATCGGTAAC
CTTCACATTTTGTGTTGCCCTAGCCATTGGAGGAAGCTGGGCTAAAGAACGT
GCTGGTGGTGCTTTCGCCGCTGGTCTTGCCTTCATCTTGATTAACCGTATC
ACTGGTACAATCTTTGGTGTATCAGGCGATATGTTGAAAAATCCAGATGCT
AGTGTTCTTGAAGCTCCAGCCTTGAACATGGGGGTTATCGTAGGGATTATC
TCAGGTTTTGTAGGGCAACTGCTTACAACAAATACTACAACCTCCGTAAA
CTTCTGATGCACCTTCATTCTTCAACGGGAAACGTTTCGTACCAATTTGTA
GTTATTCTTCGTTCAACAATCGCTGCAATCTACTTGCTGCTTTCGGCCA
GTAGTTCAAACAGGTATCAATAACTTCGGTATCTGGATTGCCAACTCACAA
GAAACTGCTCCAATCTTGCACCATTCTTGATGGTACTTTGGAACGTTTG
CTCTTGCCATTTGGTCTTCAACCATGTTGACTATCCCAATGAACTACACA
GCTCTTGGTGGTACTTATGACATTTTAACTGGTGCAGCTAAAGGTACTCAA
GTATTTCGGTCAAGACCCACTATGGCTTGCATGGGTAACAGACCTTGTA AAC
CTTAAAGGTAAGTACTGATGCTAGTCAATATCAACACTTGTAGATACAGTACAT
CCAGCTCGTTTCAAAGTTGGACAAATGATCGGTTCAATCGGTATCTTGATG
GGTGTGATGTTGCTATCTACCGTAATGTGATGCTGACAAGAAACATAAA
TACAAAGGTATGATGATTGCAACAGCTCTTGCAACATCTTGACAGGGGTT
ACTGAACCAATCGAATACATGTTTATGTTTCAATCGCAACACCTATGATCTT
GTTTACTACTTGTCAAGGTGCTGCCTTCGCTATGGCTGACGTCGTA AAC
CTACGATGCACTCATTTCGGTCAATCGAGTCTTGGACTCGTACACCTATT
GCAATCAGTGGTATTGGTATGGATATCGTTAACTTCGTTTGGGTA ACT
GTTCTCTTTGCTGTAATCATGTACTTTATCGCAAACCTCATGATTCAAAA
TTCAACTACGCAACTCCAGGGCGAACGGAACTACGAAACTGCTGAAGGT
TCAGAAGAACCAGCAGCGAAGTGAAAGTTCAGCAGGCTCTCAAGCTGTA

表2续

AACATTATCAACCTTCTTGGTGGACGTGTAACATCGTTGATGTTGATGCA
 TGTATGACTCGTCTTCGTGTAAGTGTAAAGATGCAGATAAAGTAGGAAAT
 GCAGAGCAATGGAAGCAGAAGGAGCTATGGGTCTTGTGATGAAAGGACAA
 GGGTTCAAGCTATCTACGGTCCAAAAGCTGACATTTGAAATCTGATATC
 CAAGATATCCTTGATTCAGGTGAAATCATTCCTGAAACTCTTCCAAGCCAA
 ATGACTGAAGCACAAAAACACTGTTCACTTCAAAGATCTTACTGAGGAA
 GTTACTCAGTAGCAGACGGTCAAGTTGTTGCTTTGGAAACAAGTAAAGGAT
 CCAGTATTGCTCAAAAAATGATGGGTGATGGATTTGCAGTAGAACCTGCA
 AATGGAAACATTGTATCTCCAGTTTCAGTACTGTGTCAAGCATCTTCCA
 AAAAAACATGCTTTTGGTATTGTGACGGGAAGCAGGCTTGAAGTATTGGTT
 CACATTGGTTTGGACACAGTAAGTCTTGAAGGTAACCATTACAGTTCAT
 GTTGTGAAGGACAAAAAGTTGCAGCAGGAGATCTCCTGTGCACAGCTGAC
 TTGGATGCTATCCGTGCAGCAGGACGTGAAACTTCAACAGTAGTTGTCTTC
 ACAAATGGTGATGCAATTAATCAGTTAAGTTAGAAAAACAGGTTCTCTT
 GCAGCTAAAACAGCAGTTGCTAAAGTAGAATTGTAA

ID-212A MLLQKELIPMIEANLPMAYAEKDIKFLKQQPLNN (D) YSC (S) KALC
 EYLNVSATLTRFAKKCGFKGFRQFI FKYQEMIHEKEKLALYTEATEKVL
 SDYEEMLRKTYTVLDEVQLERIAEMIETAERVYLYGKGSVLAALQEMKMR
 FMRLGVIGEVLSDEDMILWSSLLLNENCLVIGASISGQTDIVLEGLQKAA
 DKGAKTVLMTTRKFDEEDCFDELLLLASTDHLISYGNRISPQFPILLITD
 CLFSNYLESPEYQYYNQTI IHKEE*

ID-212B ATGTTACTGCAAAAAGAACAATTCCAATGATAGAAGCTAACTTACCAAT
 ATGGCATATGCTGAAAAAGACATTGCTAAATTCCTTAAAACAGCAACCT
 CTGAATRATTATTCATSTAARGCATTGTGCGAATACCTTAATGTATCCAAA
 GCAACATTGACTCGATTTGCGAAAAATGTGGTTTTAAAGGTTTTAGACAA
 TTCATTTTCAAATACCAAGAGATGATTCATGAGAAAAGAAAAGTTGGCATT
 TATACAGAGGCAACAGAAAAAGTTTTATCCGACTATGAGGAAATGTTGAGA
 AAAACTTACACGGTTCTTGATGAAGTTCAACTTGAGCGTATTGCTGAGATG
 ATAGAAACTGCTGAGCGTATATCTCTACGGTAAAGGAAGTTCTGTTCTT
 GCTTTACAAGAAATGAAGATGAGATTTATGCGTCTCGGAGTGATTGGTGAA
 GTATTATCAGACGAGGATATGATTTTGTGGAGTAGCTTACTACTTAATGAA
 AATTGCCTTGTCATTGGAGCATCCATTTCAGGTCAAACGATATGTACTA
 GAAGGTCTACAAAAGCTGCAGATAAAGGCGCTAAAACAGTTTAAATGACT
 ACAAGAAAATTTGACGAAGAAGATTGTTCTTTGATGAACTATTGTTATTA
 GCTTCGACCGATCATCTCTCGTATGGCAATCGCATATCACCTCAGTTTCCA
 ATACTTTAATTACAGACTGCTTATTCTCTAATTATCTGGAAAAGTCCAGAG
 AGACAATATTATTACAATCAAATATTATCCATAAGGAGGAATAA

ID-213A MNKSRLGRGRHGKTRHILLALIGILAISICLLGGFIAFKIYQKSFEQKI
 ESLKKEKDDQLSEGNQKEHFRQQAEVIAYYPLQGEKVISSVRELINQDV
 KDKLESKDNLVFYYTEQEEESGLKGVVNRNVTKQIYDLVAFKIEETEKTSL
 GKVHLTEDGQPFTLDQLFSDASKAKEQLIKELTSFIEDKKIEQDQSEQIV
 KNFSDQDLSAWNFDYKDSQIILYPSVVENLEEIALPVSAFFDVIQSSYL
 LEKDAALYQSYFDKHKQKVVALTFDDGPNPATTPOVLETLAKYDIKAFFV
 LGKNVSGNEDLVKRIKSEGHVVGNSHSPILSQLSLDEAKKQITDTEDEV
 LTKVLGSSSKLMRPPYGAITDDIRNSLDLSFIMWDVDSLWDKSKNEASIL
 TEIQYQVANGSIVLMHDIHSPTVNALPRVIEYLNQGYTFVTIPEMLNTR
 LKAHELIIYSRDE*

表2续

ID-213B ATGAATAAAAGTAGACTAGGACGTGGCAGACACGGGAAAACGAGACATRT
ATTATTGGCTTTGATTGGTATTTTAGCAATTTCTATTTGCCTATTAGGCG
GATTTATTGCTTTTAAGATCTACCAGCAAAAAGTTTTGAGCAAAAGATT
GAATCGCTCAAAAAGAGAAAGATGATCAATTGAGTGAGGGAAATCAGAA
GGAGCATTTTCGTACGGGCAAGCCGAAGTGATTGCCTATTATCCTCTCC
AAGGGGAGAAAGTGATTTCCTCTGTTAGGGAGT (C) TGATAAATCAAGAT
GTTAAGGACAAGCTAGAAAGTAAGGACAATCTTGTCTTACTATACAGA
GCAAGAAGAGTCAGGTTTAAAGGGAGTCGTTAATCGTAATGTGACCAAC
AAATCTATGATTTAGTTGCTTTTAAAGATTGAAGAGACTGAAAAGACCAGT
CTAGGAAAGGTTCACTTAACAGAAGATGGGCAACCTTTTACTTGGACCA
ACTGTTTTCAGATGCTAGTAAGGCTAAGGAACAGCTGATAAAAAGAGTTGA
CCTCTTCATAGAGGATAAAAAATAGAGCAAGACCAGAGTGAGCAGATT
GTAAAAAATCTCTGACCAAGACTTGTCTGCATGGAATTTTGATTACAA
GGATAGTCAGATTATCCTTTATCCAAGTCTGTGGTTGAAAATTTAGAAG
AGATAGCCTTGCCAGTATCTGCTTTCTTTGATGTTATCCAATCTTCGTAC
TTACTCGAAAAAGATGCGGCCTTGTACCAATCTTACTTTGATAAGAAACA
TCAAAAAGTTGTCGCTCTAACCTTTGATGATGGTCCAAATCCAGCAACGA
CCCCGAGGTATTAGAGACCCTAGCTAAATATGATATTAAGCGA () C
() T () TTCTTGTGCTTGGGAAAATGTTCTGGGAATGAGGACTTGG
TGAAGAGGATAAAATCTGAAGGTCATGTTGTTGGAAACCATAGCTGGAGC
CATCCGATCTCTCGCAACTCTCTTGTGATGAAGCTAAAAAGCAGATTAC
TGACTGAGGATGTGCTAACTAAAGTGTGGGTTCTAGTCTAAACTCA
TGCGTCCACCTTATGGTGCTATTACAGATGATATTCGCAATAGCTGGAT
TTGAGCTTTATCATGTGGGATGTGGATAGTCTGGACTGGAAGAGTAAAAA
TGAAGCATCTATTTGACAGAAATCAGTATCAAGTAGCTAATGGCTCTA
TCGTTTGTGATGATATTCACAGTCCGACAGTCAATGCCTTGCCAAGG
GTCATTTGAGTATTTGAAAATCAAGGTATACCTTTGTGACCATACCAGA
GATGCTCAATACTCGCCTAAAAGCTCATGAGCTGTACTATAGTCGTGATG
AATAA

ID-214A MFVKKGDKVRV IAGKDKGTEAVVLTALPKVNKVI VEGVNI VKKHQRPTNE
LPQGGIIIEKEAAIHVSNVQVLDKNGVAGRVGYKFVDGKKVRYNKKSGEVL
D*

ID-214B ATGTTTGTAaaaaaaggcgacaaagttcgcgtaatcgctggtaagataag
ggaacagaagctgttgctccttactgcccttccaaaagtaaacaaaagttatc
gttgagggtgtaacattgtaagaaacaccaacgtccaactaacgagctt
cctcaagggtggtatcatcgagaaagaagcagctatccacgtatcaaacggt
caagttttggacaaaaatggtgtagctggctggtggtggtatcaaaatttgta
gacggtaaaaaagttcgctacaacaaaaaatcaggcgaagtgcttgattaa

ID-215A MKKISNFCMLLLLCTTFFVFNVNYTREVVRIQEMGKTVDSL DLYLKDIN
EPAASVLRFFEDVSKEYKVSIIKTDSGDEVVKSQVFDKDTFPYQEFGIS
LDFTTDGEGVYSNKEISNKLGTIPFLKAKPIQLMTFQTYIKDTSRSLNG
RYTITSTQEMDKDRIVQKWSDFKIDQATLLEPTYKSAVEVINRDLLLSA
IVFVLAII LLVLT VYQPMEMKRVGVQKLLGFQDRAVLADVVKGNLYLL
LGGALVINLGVFLLDYKPKDLFPLMLWSHLLQLYLFISWLTYLIIQK
MTISSLLKGFSSFKFLIFNYVMKIGTTI LLTALLIGVGRSLEQENKELA
YQQQWVSQGNYL TLETFKLNNDLWQEELAGSGKSTDYFYRFYQDLVEKTQ
AGYVQSSSLPVKNFVQSEIQYQLTDTVDVYYANRNFLKSKGFKLPNTG
IKKVILMPASTKGEEDKNQLLGLIAFHSMKYEEQQKRTIEEMDVEIAYY
EGDWSFFPYSDKRKENLSNPIISLVNDSMMWDEKASLSTTGLNPIKIE
NTVQHQKEITELVEKLSDGNYLKFSSIQAIQEKVDSYRDAVRNFNLLFA
LFGLLSMMISYFLLVTTFLLKRRDIITKKFMGWKLVDRYRPLLVLLLLGY
SFPLLVLIFFAHAFPLLLFAGFTCLDILFVLGLASRMKRSLEVLLKGG
IL*

表2续

ID-215B ATGAAAAAATCAGTAATTTCTGTATGTTACTCCTGCTTCTGTGTACCACT
TTTTTGTTTTTAATGTAACCTATACACGAGAAGTGGTTCGGATTCAAGAA
ATGGGAAAGACTGTAGATTCTTTGGATTGTATTTGAAAGATATTAACGAA
CCTGCAGCGTCTGTTCTTCGATTTTTTGAGGATGTATCAAAGGAGTATAAA
GTCTCCATCATAAAACAGACAGTGGTGATGAGGTGGTCAAGTCTGGTGT
TTTGATAAAGATACCTTCCCCTACCAAGAGTTTGGGATTTCTTCTCTTGAT
TTTACCACAGATGGTGAAGGAGTCTATAGTAATAAGAAATTTCCAATAAA
CTTGGTACGATTCCGACCTTCTAAAAGCCAAACCTATTCAGCTTATGACT
TTTCAAACCTATATCAAGGATACATCTCGTAGTTAAATGGTCGCTATACG
ATAACTTCTACACAAGAGATGGACAAGGATAGGATTGTACAGAAATGGAGC
GATTTTTTCAAGATAGACCAGGCTACCTTGCTAGAGCCGACCTACAAAAGT
GCAGTGAAGTCATAAATCGAGATTTGCTTTTATCTGCCATTGTTTTGTC
TTGGCTATTTTGTCTTGTGTAGTGACAGTGTATCAACCGATGATGGAG
ATGAAAAGAGTTGGGGTACAAAAATTAATTTGGTTTTCAAGATAGGGCTGTT
TTAGCTGATGTTGTAAGGCAACCTTACCTCCTCCTAGGTGGGGCTCTT
GTGATCAATCTAGCGGTGTTTTCTTCTGCTTGTATATAAGCCAAAAGATTG
TTTCTATGCTGTGGTTGTCTCATTTTTTTGTCTGTGCGACTTATCTCTT
ATCAGTTGGTTGACTTACCTCTTAATCCAAAAATGACAATCAGCTCTCTG
CTGAAAGTTTTTTCATCTTCAAATTTGGTCTTATCTTCAATTATGTGATG
AAAATAGGGACAACCTATTTACTGACGGCCTTACTGATTGGGGTGGGCAGA
AGTTTTAGAAACAAGAAAACAAAGAACTTGCTTATCAGCAACAGTGGGTAAGT
CAAGGTAATTACCTGACCTTAGAAACCTTCAAACCAATGATAATCTGTGG
CAAGAAGAGCTAGCAGGGTCAGGAAATCTACAGATTATTTCTATCGATTT
TATCAGGATTTGGTAGAAAAAACGACGGCGGGCTATGTGCAGAGTAGCAGT
CTTCTGTAAAAAATTTTGTCCAATCAGAACAGATTGAGCAATATCAGTTA
ACAGATACGGTGGATGTTTACTATGCCAATCGCAATTTTCTAAAAGAGCAAG
GGATTCAGCTACCAAATACCGGTATTA AAAAAGTTATTTTGATGCCAGCA
AGTACGAAAGGTGAAGAAGATAAAAATCAGCTCTGGGGAAAGTTAATTGCC
TTTCATTGATGAAGTATGAAGAGCAGAAAAACGAACGATAGAGGAGATG
GATGTCGAGATTGCCATATTAAGAGGATTTGGTCATTTTCCCATATAGT
GATAAGCGAAAGGAAAATCTCTCAAATCAATTTATAGCTTGGTCAATGAT
TCTGATATGATGTGGGATGAGAAAGCCTCCCTGTCAACAACGGCTTAAAT
AATCCGATTA AATGAAAATACGGTTCAACATCAAAAAGAGATTACAGAG
TTAGTTGAGAAATGTCAGATGGAATATTTAAAATTTTCATCTATTCAA
GCCATTCAACAAGAGAAAGTGGATTCTTATCGAGATGCTGTTCGGAATTT
AACCTACTCTTTGCTTTGTTGGTCTCCTTAGCATGATGATTTCCCTACTTC
TTACTAGTAACAACCTTCTTATTGAAGCGCAGGGATATCATTACCAAGAAG
TTTATGGGGTGGAACTGGTCGATCGCTACCGTCTCTCCTCGTCTGCTC
TTGCTGGGCTATAGTTTCCCTCTTCTAGTCTTGATTTTCTTTGCCATGCG
TTCTTACCACTTCTACTGTTTGCAGGTTTTACATGTCTGGATATACTATTT
GTGCTAGGCTTAGCTTCTAGGATGGAGAAAAGAAGTCTAGTAGAGTTATTG
AAAGGGGCATCTTATGA

ID-216A MPITAADIRREVKEKNVTFIRLMFSDILGTMKNVEIPATDEQLDKVLSNK
VMFDGSSIEGFVRINESDMYLPDLDTWTVFPWGDENGSVAGLICDVYTT
EGEPFAGDPRGNLKRALRHMEVGFKSNLGPPEFFLFLKLDENGDPtle
VNDKGGYFDLAPDLADNTRREIVNVLTkMGFEVEASHHEVAVGQHEIDF
KYDEVLRACDKIQIFKLVVKTiARKHGLYATfMAKPKFGIAGSGMHCNMS
LFDAEGNNAFFDPNDPKGMQLSETAYHFLGGLIKHAYNYTAIMNPTVNSY
KRLVPGYEAPVYIAWAGRNRSPVLRVPASRGMGRLELRSDPMPANPYVA
MAVLLLEVGLYGIENKIEAPAPIEENIYIMTAERKEAGITDLPSTLHNAL
KALTEDEVVKAALGDHIYTSFLEAKRIEWASYATFVSQWEIDNYLDLY*

表2续

ID-216B ATGCCAATCACAGCTGCAGATATTCGTCGTGAAGTCAAGGAAAAAATGTT
ACCTTTATTCGTCTTATGTTCTCAGATATTTGGGAACCATGAAAAACGTC
GAAATTCCTGCTACAGATGAACAGTTAGATAAGGCTTGTGCAACAAGGTT
ATGTTTGATGGATCTTCTATTGAAGGTTTGTACGTATCAATGAGTCGGAT
ATGTACTTGTACCCGGACTTGGATACATGGACAGTCTTCCCTTGGGGAGAT
GAAAATGGAAGTGTGCAGGTCTGATCTGTGATGTT (C) TATACAACAGAA
GGTGAACCATTTGCGGGTGACCCTCGTGGTAATTTGAAACGAGCTCTTCGT
CACATGGAAGAAGTTGGATTCAAATCCTTCAACCTTGGTCCAGAGCCAGAA
TTCTTCTATTTAAGTTGGATGAAAATGGGGACCCAACACTTGAAGTGAAT
GACAAGGGTGGCTACTTTGACTTGGCACCTACTGACCTTGGCGACAACACA
CGTCGTGAGATTGTAATGTCTTGACCAAAATGGGATTTGAAGTAGAAGCG
AGTCACCACGAGGTTGCGGTTGGACAGCATGAGATTGACTTTAAGTACGAT
GAAGTCTCCGTGCTTGTGATAAGATTCAAATCTTTAAGCTTGTGTTAAA
ACCATTGCTCGCAAACACGGACTTTACGCAACATTTATGGCGAAGCCAAAA
TTTGGTATTGCTGGATCAGGTATGCACTGTAATATGTCCTTGTGATGCA
GAAGGAAATAACGCCTTCTTGTATCCAAATGATCCAAAAGGAATGCAGTTG
TCAGAAACAGCTTACCATTTCCTAGGCGGTTTGTATCAAGCATGCTTACAAC
TATACTGCCATCATGAACCAACAGTTAACTCATACAAACGTTTGGTTCCA
GGTTATGAAGCGCCTGTTTACATTGCTTGGGCTGGTCGTAACCGTTCCGCA
CTTGTGCGCGTACTGCTTACGTTGGTATGGGAACTCGTCTTGAGTTGCGT
TCAGTGGATCCAATGGCGAACCTTACGTTGCTATGGCTGTTCTTTGGAA
GTTGGTTGTATGTTATGAAAATAAAATCGAAGCACCAGCTCCTATCGAA
GAAAATATCTACATCATGACAGCAGAAGAGCGCAAGGAAGCTGGTATTACA
GACCTCCATCAACTCTTACAACGCTTTGAAAGCTTTGACAGAAGATGAA
GTGGTTAAAGCTGCTCTCGGAGATCACATCTATACTAGCTTCCTTGAAGCC
AAACGAATCGAATGGGCAAGTTATGCAACCTTCGTTTACAATGGGAAATT
GATAATTATTTAGACCTTTACTAA

ID-217A MYYLVLGI LLLLLLVFATPESIKGTVNIVAMVCILVALLI LVLVLSFLKIF
QLPTEIFLAIAMLILAYFSVRDITLMPVKKSKRR*

ID-217B ATGGTCTATTTAGTCTTAGGAATTTACTGCTCCTACTCTATGTATTTGCG
ACACCAGAAAGCATTAAAGGGACTGTCAATATCGTCGCTATGGTATGTATT
TTAGTGGCACTCTTGATTTTATTGGTCTATCTTTTCTGAAAATTTTCAA
TTACCAACAGAAATATTCCTAGCAATAGCCATGTTGATCCTAGCTTACTTT
AGTGTTAGAGACATCACACTCATGCCAGTCAAAAAAGTAAAAGAAGATAA

ID-219A SGLGLNFYALSSYYLGSFLAPLVYFFDLTNMPDAIYLTTLLKFLIGLST
FFSLNKLQFSI PQILKLALSTSYALMSFTVSQLEIKTWLDVFI LIPLIIT
GLHLITEKLLLYFTSLSILFIQNYFYGMTVLFILFWYLCQISWDFKT
RKSSVLD FIVISFLAGMASLIMTLPTLFDLQTHGEKLEVTKFQTESSWY
LDLFAKQFIGSFDTTKYGAIPMIFVGLFPFILITLFFTLKSIKFHVKLIY
VIFFAFLIASFYIEALDLFWQGMHTPNMFLHRYAWIFSTLLIYTAAEVLK
RLKELKWNFLVSLFLVVAGFLATIYLKSHYSLDNLNILLLEFLVVYSL
LLLAVIKKFISVNLFAILISL FILVEMSLNASSQMDGIAKEWGFASRSAY
SRDIPAMESFSTYIGNQFTRTEKLQTGTGNSMKFNNGISQFSSVRNRS
SSSTLDKLGFKSSGTNLNRYANNSILADSLFGIQYNISDSPIDKYGFKD
IYQKONLTYENQYSLPIAVASQSVYNDVKFNEHTLDNQASFLNQLANVN
FDYFSP IPEYKTEKIENTNDLISVTSSSNEDAAIQYQIEVPENSQVYLSF
INLHFSNDKQKVDILVNGEKKFTTNDNVFSFFNLGYTKEKKT FNINVSF
PGNSQVSFESPTFYRLDTKTFT EAIQIKIQPVTVSTSKNKVFATYDVQQ
DTSIFFTI PDKGWSAYQDGKKIEIKQAQTGFMMKVDIPKGGKGTITLSFIP
NGFITGAICSF TSLLLFGIYNHRRKSSKA*

表2续

ID-219B AGTGGTCTAGGGCTAAACTTCTATGCCCTATCTAGTTATTACTTGGGTAGT
 TTTCTCGCGCCTCTGGTTACTTTTTTGATCTAACGAATATGCCAGATGCT
 ATCTATCTGACAACTCTCTAAAATTTGGATTGATTGGTCTGTCAACCTTT
 TTTAGTTTGAATAAATGTTTCAATCTATCCCTCAGATTTAAAAGTAGCC
 TTATCTACTTCTATGCTCTGATGAGTTTACTGTCAGTCAATTAGAGATA
 AAAACCTGGCTAGATGTTTTTATCTTGATTCCTTTAATTATAACTGGTTTA
 CATCTACTGATAACTGAAAAGAACTCCTATTGTACTTTACAAGTCTGTCA
 ATCTTATTTATTCAAATTTATTTTGGATATATGACAGTATTGTTTCTT
 ATTTTCTGGTATCTCTGTCAAATTTTCGTGGGACTTTAAGACTCGAAAATCA
 TCTGTTCTTGATTTTCATAGTTATCTCCTTTTATAGCTGGTATGGCTAGTTG
 ATTATGACTCTTCCCACTCTATTTGATTTACAGACACATGGGGAAAAATTG
 ACTGAAGTTACAAAGTTTCAAAGTAAAGTAGCTGGTATCTTGATCTCTTT
 GCTAAGCAATTCATTGGTTCCCTTTGACACAACAAAGTATGGGGCCATCCCA
 ATGATTTTTGTTGGACTATTTCCCTTTATTTGACCATTTTATTTTTTACG
 CTGAAATCTATTAAGTTTCCAGTGAAGTCAATATATGTAATATCTTTGCA
 TTTCTAATTGCAAGCTTTTACATAGAAGCTCTTGACTTATTTTGGCAAGGC
 ATGCATACTCCAACATGTTTTTACATCGCTATGCTTGGATTTTCTCTACC
 TTGTTAATTTACACAGCAGCAGAAGTCTTAAAGCGTCTGAAAGAACTTAAA
 GTCTGGAAATTTTTAGTTTCGCTTTTTCTTGATAGCAGGATTTTTAGCT
 ACCATCTATCTAAAATCGCATTATCTTTTTTAAACAGATTTGAATATCTG
 CTTACTCTGAAATTTTGGTTGTCTATCTCTTTTACTCCTTGCAAGTTATC
 AAAAAGTTTATATCTGTGAATCTATTTGCCATTCTAATCTCTTTATTTATA
 CTGGTTGAAATGAGTTTAAATGCTTCACTCAAATGGACGGAATTGCTAAG
 GAATGGGGATTTGCTTCTCGAAGTCTTATAGTCTGAGATATCCAGCTATG
 GAATCTTTCTCAACATATATTGGAAATCAATTTACTCGTACTGAAAACTA
 CAAACTCAGACAGGAAATGACAGTATGAAATTCAACTACAATGGAATCTCT
 CAATTTTCACTGTTCGAAATCGTTCAAGCTCTACTTTAGATAAACTT
 GGTTTTTAAATCCTCTGGGACTAATCTCAATCTCCGATATGCAAATAATAGT
 ATTTTGGCTGATAGTTTATTTGGTATCCAGTACAATATCTCAGACAGTCTT
 ATTGATAAGTATGGCTTTAAAGATATCTATCAAAAAGATAATCTTACCCTA
 TATGAAAATCAACTCTCTTCCGATTGCAGTTGCGAGTCAATCTGTTTAC
 AATGATGTCAAGTTCAATGAACATACCTTGGATAATCAGGCCTCATTTTTTA
 AATCAACTTGCTAACGTCATTTTGGATTATTTTTCTCCAATACCTTATGAA
 AAAACAGAAAAAATAGAAAATACTAATGATTTGATTAGTGTCAAGTTCT
 TCAAATGAAGATGCAGCAATCCAGTATCAAATGAAGTTCCAGAAAACAGC
 CAAGTTTATCTCTCTTTCATAAACCTTCACTTTTCTAACGATAAACAAAAG
 AAGGTTGACATCCTTGTAATGGTGAAAAAAGACTTTTACAACCTGATAAT
 GTCTTCTCCTTCTTAATCTAGGATATACTAAAGAGAAAAAACTTTCAAT
 ATCAATGTTAGTTTCCCTGGAAATTCACAAGTATCATTTGAATCTCCTACC
 TTCTATCGTTTTAGATACCAAAATTTCAACGAGGCAATTCAAAAAATTAAA
 GAACAACCTGTACAGTATCAACTTCTAAAAACAAGGTTTTTGTACATAT
 GATGTCCAACAAGATACATCTATTTTCTTACCATTCTTTATGACAAAGGT
 TGGTCTGCCTACCAAGATGTAAGAAAATAGAAATTAACAAGCTCAAAC
 GGATTTATGAAAGTTGACATTTCCAAGGGGAAAGGAAGTATTACACTTCC
 TTCATTTCCAATGGTTTTTATTAAGGCAATCTGTTTCTTACTTCTCTC
 TTACTATTTGGAATCTATAATCACAGACGAAAGTCACTAAGGCATAA

ID-220A MNEKVFDRPVHNYIHVNNQIYYDLINP(T)Q(K)EFQRLRRIKQLGTSY
 TFHGGEHSRFSHCLGVYEIARRITEIFEKYPEEWNPAESLLTMTAALLH
 DLGHGAYSHTFEHLFDTDHEAITQEI I QNPETEIHQVLLQVAPDFPEKVA
 SVIDHTYPNKQVVLISSQIDADRM DYLLRDSYFTGASYGEFDLTRLIRV
 IRPIENGI AFQRNGMHAI EDYVLSRYQMYMQVYFHPATRAMEVLLQNLK
 RAKELYPEDKDFARTSPHLLPFFEKNVTLTDYLALDDGVMNTYFQLWMT
 SPDKILADLSHRFVNRKVFKSITFSQEDQDQLTSMRKLVEDIGFDPDYTT
 AIHKNFDLPYDIYRPESENPRQTQIEILQKNGELAE LSSLSPIVQSLAGSR
 HGDNRIFYPKEMLDQNSIFASITQQFLHL*

表3

ID-304L2A
 MELVLPNNYVALEQEEMMYLDGGFSSILRWPVATAINIAFNGVLGGGAI SLVRNYIRN
 YGLGRVTSIAIAGAAARYVGV RVANRVAGFALSAINGFAAWMSIGDAITTIWANNDVN
 RRDPNLNALW

ID-304L2B
 ATGGAACCTCGTATTACCAAATAATTATGTTGCTCTTGAGCAAGAAGAGATGATGTAT
 CTTGATGGGGGATTTTCTATTCTGAGATGGCCTGTGCAACAGCCATTAATATAGCT
 TTTAATGGTGTTT TAGGTGGAGGAGCAATCAGTCTAGTTAGAAAT TATATTGTAAT
 TATGGTTTGGGGCAGTTACAAGCGCAATTGCTGGAGCAGCTGCAAGATATGTTGGG
 GTACGAGTTGCAAATAGAGTGGCAGGATTTGCACTGTCTGCTATTAATGGATTGCA
 GCTTGGATGTCAATTGGCGATGCTATTACAACAATCTGGGCCAACCAATGATGTAAAT
 AGGAGAGACCCAAATTTAAACGCCTTGTGGTAA

ID-304L3A
 MELVLPNNYVVIDEEMMYLDGGAYLSKRACQGICAAALAMSPGTFIALAGAAVLTKK
 LINYIKVGGGLGGWLGAAAGVLAGAAGRIAYCIGYGALNRGCDISGNPYPWDGFISA
 TVR

ID-304L3B
 ATGGAACCTGTATTACCAAATAATTATGTTGTGATTGATGAAGAAGAGATGATGTAC
 CTTGATGGGGGAGCTTATTTAAGCAAGCGTGCTTGTCAAGGAATTTGCGCAGCTTTA
 GCTATGAGTCCAGGAACCTTTTATAGCATTAGCTGGAGCTGCAGTTTAAACCAAAAA
 CTAATAAAC TATATTAAGTTGGAGGCCCTTGGAGGTTGGCTTATTTGGTGCAGCAGCA
 GGTGTATTGGCTGGGGCGGCAGGAAGAATAGCTTACTGTATTGGATATGGTGTCTT
 AATAGAGGTTGTGATATTAGCGGGAACCCCTTATCCTTGGGATGGATT CATATCTGCC
 ACAGTAAGATGA

ID-304L4A
 MELVLPNNYVVIDEEMMYLDGEAYLSKRACQGICAAALAMSSGTFIALAGAAVLTKK
 LINYIKVGGGLGGWLGAAAGVLATAAGKIAYYIGYV LN RGCDINGNPYPWDGFISA
 TVR

ID-304L4B
 ATGGAACCTGTATTACCAAATAATTATGTTGTGATTGATGAAGAAGAAATGATGTAT
 CTTGATGGGGGAGCTTATTTAAGCAAGCGTGCTTGTCAAGGAATTTGCGCAGCTTTA
 GCTATGAGTTCAGGCACCTTTTATAGCATTAGCTGGAGCTGCAGTTTAAACCAAAAA
 CTAATAAAC TATATTAAGTTGGAGGCTTGGAGGCTGGCTTATTTGGTGCAGCAGCA
 GGTGTATTGGCTACAGCAGCAGGAAAATAGCTTACTATATTGGATATGGTGTCTT
 AATAGAGGTTGTGATATTAACGGGAACCCCTTATCCTTGGGATGGATT CATATCTGCC
 ACAGTAAGATGAGTAATGTAG

ID-304L5A

MKQFQLRRRKQMELVLPNNYVVIDEEMMYLDGGAYLSKRACQGICVALAMSPGIFI
ALAGAAVLTKKLINIKVGGGLGGWLGAAAGVLATAAGKIAYCIGYGALNRGCDISG
NPYPWDGFISATVR

ID-304L5B

ATGGAACCTTGATTACCAAATAATTATGTTGTGATTGATGAAGAAGAAATGATGTAT
CTTGATGGGGGAGCTTATTTAAGCAAGCGTGCTTGCAAGGAATTTGCGTAGCTTTA
GCTATGAGTCCAGGAATTTTATAGCATTAGCTGGAGCTGCAGTTTTAACCAAAAA
CTAATAAACTATATTAAGGTTGGAGGCTTGGAGGCTGGCTTATTGGTGCAGCAGCA
GGTGTATTGGCTACAGCAGCAGGAAAAATAGCTTACTGTATTGGATATGGTGCTCTT
AATAGAGGTTGTGATATTAGCGGGAACCCCTTATCCTGGGATGGATTCATATCTGCG
ACAGTAAGATGA

ID-304L6A

MELVLPNNYVVIDEEMMYLDGGAIYIPRWAITGAITGAAYAALAAAGGGGLQLVLA
SYGLRSALVAGIVKGLGVLIHIGNAFANTVIRSIASAGIGAGADWIFTNIDGWDG
RRDNQLRIG

ID-304L6B

ATGGAACCTTGATTACCAAATAATTATGTTGTGATTGATGAAGAAGAGATGATGTAC
CTTGATGGGGGGCTATATATATACCAGGTGGGCAATTACAGGAGCCATTACTGGT
GCAGCATATGCAGCATTAGCAGCAGCAGGAGGTGGAGGCCCTCAACTAGTTCTTGCA
TCTTATGGATTACGCTCCGCACTGGTAGCTGGGATGTTAAAGGTTTAGGAGTATTA
GGAATTCATATTGGAAATGCTTTTGCAAATACGTTATTAGAAGTATTGCATCTGCT
GGAATTGGTGCTGGAGCTGATTGGATTTTACCAATATTATGATGGCTGGGATGGG
CGACGTGATAATCAATTGAGAATAGGTTAA

ID-304L7A

MELVLPNNYVDLEQEEMMYLDGGGVGRNWNNSRGSFATVLDVGLAIYSGGATIYSAY
AIKKAISANRGAI TRTLRLSLIKHVGSAAHGLVNTALNVALTVTGFSLGGAIAYGAD
WADGSLDGYIFA

ID-304L7B

ATGGAACCTCGTATTACCAAATAATTATGTTGATCTTGAGCAAGAAGAGATGATGTAT
CTTGATGGGGGTGGTGTGGTCGTAACGGTGAATAGTAGAGGTAGTTTGCAACA
GTTCTGGATGTAGGTTTGGCCATCTATAGTGGTGGTGAACAATTTATTCTGCTTAT
GCGATAAAAAAGCTATCTCAGCTAATAGAGGGGCTATTACGAGAACATTACGTAGT
TTAATAATTAACATGTAGGTAGTGCAGCTGGCCATTTAGTCAATACTGCACTAAC
GTTGCACTAACTGTTACTGGATTTTCACTAGGTGGAGCAATCGCATATGGGGCTGAT
TGGGCTGACGGTAGCTTAGATGGTTATATTTTGCTTAA

表3续

| | | | |
|----------------|--|---------|------------|
| 专利名称(译) | 分泌型肺炎链球菌蛋白质 | | |
| 公开(公告)号 | CN1513058A | 公开(公告)日 | 2004-07-14 |
| 申请号 | CN02810868.X | 申请日 | 2002-03-28 |
| [标]申请(专利权)人(译) | 微生物技术有限公司 | | |
| 申请(专利权)人(译) | 微生物技术有限公司 | | |
| 当前申请(专利权)人(译) | 微生物技术有限公司 | | |
| [标]发明人 | RWF勒佩吉 K皮克 JM韦尔斯 SB汉尼菲 | | |
| 发明人 | R·W·F·勒佩吉 D·巴德考克 P·J·H·斯泽 K·皮克 J·M·韦尔斯 S·B·汉尼菲 | | |
| IPC分类号 | A61K39/00 C07K14/315 C12N1/21 C12N15/31 C07K16/12 C12Q1/68 G01N33/53 A61K39/09 | | |
| CPC分类号 | C07K14/3156 A61K39/00 | | |
| 代理人(译) | 唐伟杰 | | |
| 优先权 | 2001008079 2001-03-30 GB | | |
| 外部链接 | Espacenet SIPO | | |

摘要(译)

描述了肺炎链球菌新的蛋白质，及其编码核酸序列。还描述了它们在疫苗和筛选方法中的应用。

