

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷
C12N 15/70
C07K 14/22



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 01808750.7

[45] 授权公告日 2005 年 5 月 11 日 [11] 授权公告号 CN 1201011C

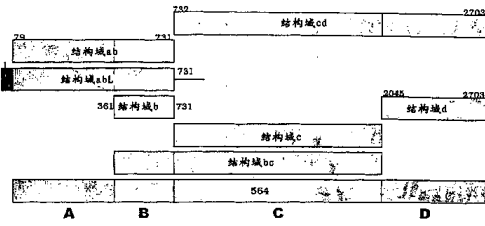
[22] 申请日 2001.2.28 [21] 申请号 01808750.7
[30] 优先权
[32] 2000. 2.28 [33] GB [31] 0004695.3
[32] 2000. 11.13 [33] GB [31] 0027675.8
[86] 国际申请 PCT/IB2001/000452 2001.2.28
[87] 国际公布 WO2001/064922 英 2001.9.7
[85] 进入国家阶段日期 2002.10.28
[71] 专利权人 启龙股份公司
地址 意大利锡耶纳
[72] 发明人 M·B·阿科 M·科曼杜西
C·加莱蒂 V·马斯格阿尼
M·M·朱利亚尼 M·皮扎
审查员 吴希哲

[74] 专利代理机构 上海专利商标事务所有限公司
代理人 徐 迅

权利要求书 2 页 说明书 99 页 附图 13 页

[54] 发明名称 奈瑟球菌蛋白质的异源表达
[57] 摘要

异源表达脑膜炎奈瑟球菌和淋病奈瑟球菌蛋白质的其它和改进的方法。 这些方法通常影响表达的水平、纯化的简易性、细胞的定位和/或所表达的蛋白质的免疫特性。



1. 一种异源表达蛋白质的方法，其特征在于，所述的蛋白质中的聚-甘氨酸段发生缺失，所述蛋白质是 287、741、983 或 Tbp2，

5 蛋白质 287 的序列如下所示：

1	CGGGGGSPD	VKSADTLSKP	AAPVVSEKET	EAKEDAPQAG	SQGQGAPSAQ
51	GSQDMAAVSE	ENTGNGGAVT	ADNPKNEDV	AQNDMPQNAA	GTDSSPTNHT
101	PDPNMLAGNM	ENQATDAGES	SQPANQPDMA	NAADGMQGDD	PSAGGQNAGN
151	TAAQGANQAG	NNQAAGSSDP	IPASNPAPAN	GGSNFGRVDL	ANGVLIDGPS
201	QNITLTHCKG	DSCSGNNFLD	EEVQLKSEFE	KLSDADKISN	YKKDGKNDKF
251	VGLVADSVQM	KGINQYIIFY	KPKPTSFAF	RRSARSRRSL	PAEMPLIPVN
301	QADTLIVDGE	AVSLTGHSN	IFAPEGNYRY	LTYGAEKLP	GSYALRVQGE
351	PAKGEMLAGA	AVYNGEVLHF	HTENGRPYPT	RGRFAAKVDF	GSKSVDGIID
401	SGDDLHMGTQ	KFKAAIDGNG	FKGTWTENG	GDVSGKFYGP	AGEEVAGKYS
451	YRPTDAEKGG	FGVFAGKKEQ	D		

蛋白质 741 的序列如下所示：

1	CSSGGGGVAA	DIGAGLADAL	TAPLDHKDKG	LQSLTLDQSV	RKNEKLKLA
51	QGAEKTYGNG	DSLNTGKLKN	DKVSRDFIR	QIEVDGQLIT	LESGEFQVYK
101	QSHSALTAFQ	TEQIQDSEHS	GKMVAKRQFR	IGDIAGEHTS	FDKLPEGGRA
151	TYRGTAFGSD	DAGGKLTITI	DFAAKQGNK	IEHLKSPELN	VDLAAADIKP
201	DGKRHAVISG	SVLYNQAEKG	SYSLGIFGGK	AQEVAGSAEV	KTVNGIRHIG
251	LAACKQ				

蛋白质 983 的序列如下所示：

1	CLGGGGGGTS	APDFNAGGTG	IGSNSRATTA	KSAAVSYAGI	KNEMCKDRSM
51	LCAGRDDVAV	TDRDAKINAP	PPNLHTGDFP	NPNDAYKNLI	NLKPAIEAGY
101	TGRGVEVGIV	DTGESVGSIS	FPELYGRKEH	GYNENYKNYT	AYMRKEAPED
151	GGGKDIEASF	DDEAVIETEA	KPTDIRHVKE	IGHIDLVSII	IGGRSVDGRP
201	AGGIAPDATL	HIMNTNDETK	NEMMVAAIRN	AWVKLGGERV	RIVNNSFGTT
251	SRAGTADLFQ	IANSEEQYRQ	ALLDYSGGDK	TDEGIRLMQQ	SDYGNLSYHI
301	RNKNMLFIFS	TGNDAAQAPN	TYALLPFYEK	DAQKGIITVA	GVDRSGEKFK
351	REMYGEPGTE	PLEYGSNHCG	ITAMWCLSAP	YEASVRFRTR	NPIQIAGTSF
401	SAPIVTGTAA	LLLQKYPWMS	NDNLRTTLLT	TAQDIGAVGV	DSKFGWGLLD
451	AGKAMNGPAS	FPFGDFTADT	KGTSDIAYSF	RNDISGTGGL	IKKGSQQLQL
501	HGNNTYTGKT	IIEGGSLVLY	GNNKSDMRVE	TKGALIYNGA	ASGGSLNSDG
551	IVYLADTDQS	GANETVHIKG	SLQLDGKGT	YTRLGKLLKV	DGTAIIGGKL
601	YMSARGKGAG	YLNSTGRRVP	FLSAAKIGQD	YSFFTNIETD	GGLLASLDSV
651	EKTAGEGDT	LSYYVRRGNA	ARTASAAHS	APAGLKHAVE	QGGSNLENLM
701	VELDASESSA	TPETVETAAA	DRTDMPGIRP	YGATFRAAAA	VQHANAADGV
751	RIFNSLAATV	YADSTAAHAD	MQGRRLKAVS	DGLDHNGTGL	RVIAQTQQDG

801	GTWEQGGVEG	KMRGSTQTVG	IAAKTGENTT	AAATLGMGRS	TWSENSANAK
851	TDSISLFAGI	RHDAGDIGYL	KGLFSYGRYK	NSISRSTGAD	EHAEGSVNGT
901	LMQLGALGGV	NVPFAATGDL	TVEGGLRYDL	LKQDAFAEKG	SALGWSGNSL
951	TEGTLVGLAG	LKLSQPLSDK	AVLFATAGVE	RDLNGRDYTV	TGGFTGATAA
1001	TGKTGARNMP	HTRLVAGLGA	DVEFGNGWNG	LARYSYAGSK	QYGNHSGRVG
1051	VGYRF				

蛋白质 Tbp2 的序列如下所示:

1	CLGGGGSFDL	DSVDTEAPRP	APKYQDVFSE	KPQAQKDQGG	YGFAMRLKRR
51	NWYPQAKEDE	VKLDESDEWA	TGLPDEPKEL	PKRQKSVEIK	VETDSDNNIY
101	SSPYLKPSNH	QNGNTGNGIN	QPKNQAKDYE	NFKYVYSGWF	YKHAKREFNL
151	KVEPKSAKNG	DDGYIFYHGK	EPSRQLPASG	KITYKGVWHF	ATDTKKGQKF
201	REIIQPSKSQ	GDRYSGFSGD	DGEYSNKNK	STLTDGQEGY	GFTSNLEVDF
251	HNKCLTGKLI	RNNANTDNNQ	ATTTQYYSLE	AQVTGNRFNG	KATATDKPQQ
301	NSETKEHPFV	SDSSSLSGGF	FGPQGEELGF	RFLSDDQKVA	VVGSATKDK
351	PANGNTAAAS	GGTDAAASNG	AAGTSSENGK	LTTVLDAVEL	KLGDKEVQKL
401	DNFSNAAQLV	VDGIMIPLLP	EASESGNNQA	NQGTNGGTAF	TRKFDHTPES
451	DKKDAQAGTQ	TNGAQASNT	AGDTNGKTKT	YEVECCSNL	NYLKYGMLTR
501	KNSKSAMQAG	ESSSQADAKT	EQVEQSMFLQ	GERTDEKEIP	SEQNIVYRGS
551	WYGYIANDKS	TSWSGNASNA	TSGNRAEFTV	NFADKKITGT	LTADNRQEAT
601	FTIDGNIKDN	GPEGTAKTAE	SGFDLDQSNT	TRTPKAYITD	AKVQGGFYGP
651	KAEELGGWFA	YPGDKQTKNA	TNASGNSSAT	VVFGAKRQQP	VR。

2. 如权利要求 1 所述的方法, 其特征在于, 从 287 或 983 除去(Gly)₆。
3. 如权利要求 1 所述的方法, 其特征在于, 从 Tbp2 或 741 除去(Gly)₄。
4. 如权利要求 2 或 3 所述的方法, 其特征在于, 还除去前导肽。
5. 如权利要求 1-3 中任一项所述的方法, 其特征在于, 所述的异源表达是在大肠杆菌宿主中进行。
6. 如权利要求 4 所述的方法, 其特征在于, 所述的异源表达是在大肠杆菌宿主中进行。

奈瑟球菌蛋白质的异源表达

5 本文将所引用的全部文献都纳入作为参考。

技术领域

本发明涉及蛋白质表达的领域。具体说，本发明涉及奈瑟球菌(如淋病奈瑟球菌或较佳地为脑膜炎奈瑟球菌)的蛋白质的异源表达。

10

背景技术

国际专利申请 WO 99/24578、W099/36544、W099/57280 和 WO 00/22430 公开了脑膜炎奈瑟球菌(*Neisseria meningitidis*)和淋病奈瑟球菌(*Neisseria gonorrhoeae*)的蛋白质。这些蛋白质一般是以 N-末端 GST-融合体或 C-末端 His-标记融合体在大肠杆菌中表达的(即异源表达)，虽然也公开了其它表达系统(包括在天然的奈瑟球菌中的表达)。

本发明的目的是提供这些蛋白质的异源表达的其它或改进方法。这些方法通常影响表达的水平、纯化的简易程度、表达在细胞内定位和/或表达的蛋白质的免疫学特性。

20

发明的公开

本文的命名

本文参考在 WO 99/24578、WO 99/36544 和 WO 99/57280 中公开的 2166 个蛋白质序列，并将它们编号为如下的 SEQ#：

申请	蛋白质序列	本文的 SEQ#
WO 99/24578	偶 SEQ ID 2-892	SEQ# 1-446
WO 99/36544	偶 SEQ ID 2-90	SEQ#447-491
WO 99/57280	偶 SEQ ID 2-3020	SEQ#492-2001
	偶 SEQ ID 3040-3114	SEQ#2002-2039

	SEQ ID 3115-3241	SEQ#2040-2166
--	------------------	---------------

除了这种的 SEQ#编号外，本文还使用了 WO 99/24578、WO 99/36544 和 WO 99/27280 中的命名约定(如 WO99/24578 和 WO 99/36544 中用的‘ORF4’、‘ORF40’、‘OFR40-1’等；WO 99/57280 中用的 ‘m919’、 ‘g919’ 和 ‘a919’ 等)。

在本文中将 Tettelin 等[Science(2000)287:1809-1815]中的从 NMB0001 5 到 NMB2160 的 2160 个蛋白质称为 SEQ#2167-4326[参见 WO 00/66791]。

本文采用的术语“本发明的蛋白质”指包含以下的蛋白质：

- (a)SEQ#1-4326 中的一个序列；或
- (b)与 SEQ#1-4326 中的一个序列相同的序列；或
- (c)SEQ#1-4326 中的一个序列的片段。

10 (b)中的“序列相同性”的程度最好大于 50%(如 60%、70%、80%、90%、95%、99%或更大)。其包括突变体和等位基因变体[如，参见 WO 00/66741]。相同性宜用 Smith-Waterman 同源性搜寻算法确定，如在 MPSRCH 程序(Oxford Molecular)中执行的，采用参数“缺口罚分(gap open penalty)”为 12，“缺口延伸罚分(gap extension penalty)”为 1 进行缺口仿射搜索。通常，将两种蛋白质之
15 间 50%或更高的相同性视为功能等效的指示。

(c)中的“片段”应包含 SEQ#1-4326 中一个序列的至少 n 个连续的氨基酸，且根据具体的序列 n 为 7 或更高(如 8、10、12、14、16、18、20、25、30、35、40、50、60、70、80、90、100 或更高)。较佳地，片段包含 SEQ#1-4326 中一个序列的表位。优选的片段是在 WO 00/71574 和 WO 01/04316 中公开的那些。

20 本发明优选的蛋白质是在脑膜炎奈瑟球菌血清群 B 中发现的。

根据本发明使用的优选蛋白质是血清群 B 脑膜炎奈瑟球菌菌株 2996 或菌株 394/98(新西兰菌株)。除非特别指出，本文所述的蛋白质是脑膜炎奈瑟球菌菌株 2996 的蛋白质。但是，应该理解通常本发明并不受菌株的限制。参考具体的蛋白质(如 ‘287’、 ‘919’ 等)可以包括任何菌株的该蛋白质。

25

非融合表达

在异源表达的第一步骤中，没有使用融合配体，而是使用了天然前导肽(如果存在)。这通常防止了来自融合配体的任何“干扰”，并可能改变异源宿主

中的细胞定位和/或翻译后的修饰和/或折叠。

因此，本发明提供了一种异源表达本发明蛋白质的方法，在该方法中(a)没有使用融合配体，并(b)使用了蛋白质的天然前导肽(如果存在)。

该方法通常包括制备表达本发明蛋白质的载体的步骤，这样第一个表达的
5 氨基酸是该蛋白质的第一个氨基酸(甲硫氨酸)，且最后一个表达的氨基酸是该蛋白质的最后一个氨基酸(即，在天然 STOP 密码子前的密码子)。

最好将该方法用于表达使用天然前导肽的如下蛋白质：111、149、206、
225-1、235、247-1、274、283、286、292、401、406、502-1、503、519-1、525-1、
552、556、557、570、576-1、580、583、664、759、907、913、920-1、936-1、
10 953、961、983、989、Orf4、Orf7-1、Orf9-1、Orf23、Orf25、Orf37、Orf38、
Orf40、Orf40.1、Orf40.2、Orf72-1、Orf76-1、Orf85-2、Orf91、Orf97-1、Orf119、
Orf143.1、NMB0109 和 NMB2050。本文蛋白质命名中所用的后缀“L”指用天然前导肽按这种方式的表达。

最好用该方法(不使用融合配体，且不含天然前导肽)表达的蛋白质包括：
15 008、105、117-1、121-1、122-1、128-1、148、216、243、308、593、652、726、
926、982、Orf83-1 和 Orf143-1。

有利地，用这种方法来表达 ORF25 或 ORF40，形成的蛋白质诱导的抗细菌抗体比 GST-或 His 融合体诱导得更好。

这种方法特别适用于表达脂蛋白。

20

前导肽的替代

在异源表达的第二步骤中，用不同蛋白质的天然前导肽替代本发明蛋白质的天然前导肽。另外，最好不使用融合配体。在异源宿主中使用蛋白质自身的前导肽通常会将蛋白质定位于其‘天然’细胞位置，有时异源宿主不能有效地
25 识别该前导序列。在这种情况下，则可改为使用已知能有效地驱动蛋白质靶向的前导肽。

因此，本发明提供了一种异源表达本发明蛋白质的方法，其中(a)用不同蛋白质的前导肽替代该蛋白质的前导肽，和任选地(b)不使用融合配体。

该方法通常包括如下步骤：获得编码本发明蛋白质的核酸；操作该核酸以

除去编码所述蛋白质前导肽的核苷酸，并引入编码不同蛋白质的前导肽的核苷酸。可以将得到的核酸插入表达载体中，或作为表达载体的一部分。表达的蛋白质包含 N-末端的替代前导肽，其后为本发明减去前导肽的蛋白质。

前导肽最好为本发明另一蛋白质(如 SEQ#1-4326 中的一种)的，但也可以是 5 是大肠杆菌的蛋白质(如，OmpA 前导肽)或胡萝卜软腐欧文氏菌(*Erwinia carotovora*)的蛋白质(如 PelB 前导肽)。

特别有用的置换前导肽是 ORF4 的前导肽。该前导肽能指导大肠杆菌中的脂质化(lipidation)、改善细胞定位，且其特别适用于表达蛋白质 287,919 和 Δ G287.961 的前导肽和 N-末端区域也特别有用。

10 另一种可用的置换前导肽是大肠杆菌 OmpA 的。该前导肽能指导在大肠杆菌中的膜定位。其特别有利于 ORF1 的表达，形成的蛋白质比从其自身前导肽融合和表达的蛋白质能诱导更好的抗细菌抗体。

另一种可用的置换前导肽是 MKKYLFSAA。它能直接分泌到培养基中，是非常短和有活性。这种前导肽的使用不限于奈瑟球菌蛋白质的表达-可用它来指 15 导任何蛋白质(尤其是细菌蛋白质)的表达。

前导肽的缺失

在异源表达的第三步骤中，删除本发明蛋白质的天然前导肽。另外，不宜使用任何融合配体。

20 因此，本发明提供了一种异源表达本发明蛋白的方法，其中(a)删除蛋白质的前导肽，和任选地(b)不使用任何融合配体。

该方法通常还包括如下步骤：获得编码本发明蛋白质的核酸；操作该核酸以除去编码所述蛋白质前导肽的核苷酸。可以将得到的核酸插入表达载体中，或作为表达载体的一部分。表达的蛋白质的第一氨基酸是成熟的天然蛋白质的 25 第一氨基酸。

该方法可以提高表达的水平。例如，对蛋白质 919 而言，删除前导肽时，大肠杆菌中的表达水平高许多。没有前导肽时，表达的提高是由于定位的改变。

该方法还优先用于 919、ORF46、961、050-1、760 和 287 的表达。

基于区域的表达

在异源表达的第四步骤中,该蛋白质以结构域表达。这可以与融合系统(如 GST 或 His-标记的融合)联合使用。

因此,本发明提供了一种异源表达本发明蛋白质的方法,其中(a)删除蛋白质的至少一个结构域,和任选地(b)不使用融合配体。

该方法通常还包括如下步骤:获得编码本发明蛋白质的核酸;操作该核酸以除去蛋白质中的至少一个结构域。可以将得到的核酸插入表达载体中,或作为表达载体的一部分。不使用融合配体,表达的蛋白质的第一氨基酸是该蛋白质结构域的第一氨基酸。

10 通常,用数据库中已知的序列将蛋白质排序,然后确定该蛋白质中显示彼此不同的排序模式的区域,从而将该蛋白质分成概念上的结构域。

最好将该方法用于表达蛋白质 287。概念上可以将该蛋白质分成 3 个结构域,称为 A、B 和 C(见图 5)。结构域 B 与 IgA 蛋白酶排序很接近,结构域 C 与运铁蛋白结合的蛋白质排序很接近,而结构域 A 显示与数据库列不接近。WO 00/66741 公开了 287 的多聚体形式的排序。

一旦将蛋白质分区成结构域,就可将它们(a)单独表达(b)以蛋白质的缺失形式表达(如蛋白质 ABCD→ABD、ACD、BCD 等)、或(c)重排表达(如蛋白质 ABC→ACB、CAB 等)。这三种策略可与所需的融合配体联合。

20 还将 ORF46 分为 2 个结构域-第一结构域(氨基酸 1-433)(在品种和血清群之间很好保守的)和第二结构域(氨基酸 433-608)(非很好保守的)。最好删除第二结构域。WO 00/66741 公开了 ORF46 的多聚体形式的排序。

还将蛋白质 564 分成结构域(图 8)如具有蛋白质 961(图 12)和蛋白质 502(MC58 蛋白质的氨基酸 28-167)。

25 杂交蛋白质

在异源表达的第五步骤中,将本发明的 2 种或多种(如 3、4、5、6 或更多)蛋白质表达为单杂交蛋白质。不宜使用非奈瑟球菌融合配体(如 GST 或聚-His)。

这能提供 2 个优点。其一,该蛋白质自身可能不稳定或表达很弱,而加入合适的能克服该问题的杂交配体就可协助该蛋白质。其二,简化工业生产-产

生两种分开的-有用的蛋白质只需要一次表达和纯化。

因此,本发明提供了同时异源表达两种或多种本发明蛋白质的方法,其中所述的两种或多种本发明的蛋白质是融合的(即,它们是以单多肽链翻译的)。

该方法通常还包括如下步骤:获得编码本发明第一种蛋白质的第一种核
5 酸;制得编码本发明第二种蛋白质的第二种核酸;连接第一和第二核酸。将得到的核酸插入表达载体中,或作为表达载体的一部分。

较佳地,本发明在杂交蛋白质中构成的蛋白质是来源于同一菌株的。

可直接连接杂交体中的融合蛋白;或通过接头肽连接,如通过聚-甘氨酸
10 接头(即, G_n , 其中 $n=3, 4, 5, 6, 7, 8, 9, 10$ 或更高);或通过协助克隆的短肽序列连接。显然,不宜将 ΔG 蛋白质连接于聚-甘氨酸接头的C-末端。

融合的蛋白质可能缺少天然前导肽,或可能包含N-末端融合配体的前导肽序列。

这种方法适用于表达蛋白质 orf1, orf25, orf40, Orf46/46.1, orf83, 233, 287, 564, 687, 741, 907, 919, 953, 961 和 983。

15 由式 $NH_2-A-B-COOH$ 在下表中以 'X' 表示的 42 个杂交体是优选的:

A	B→	ORF46.1	287	741	919	953	961	983
ORF46.1			X	X	X	X	X	X
287	X			X	X	X	X	X
741	X	X	X		X	X	X	X
919	X	X	X	X		X	X	X
953	X	X	X	X	X		X	X
961	X	X	X	X	X	X		X
983	X	X	X	X	X	X	X	

因此,优选的以杂交体表达的蛋白质是 ORF46.1, 287, 741, 919, 961 和 983。它们可以全长形式或聚-甘氨酸缺失(ΔG)形式使用(如 ΔG -287、 $\Delta GTbp2$ 、 $\Delta G741$ 、 $\Delta G983$ 等),或以截短形式使用(如 $\Delta 1$ -287、 $\Delta 2$ -287等),或以结构
20 域缺失的形式使用(如 287B、287C、287BC、ORF46₁₋₄₃₃、ORF46₄₃₃₋₆₀₈、ORF46、961c等)。

特别优选的是:(a)含 919 和 287 的杂交蛋白;(b)含 953 和 287 的杂交蛋白;(c)含 287 和 ORF46.1 的杂交蛋白;(d)含 ORF1 和 ORF46.1 的杂交蛋白;

(e)含 919 和 ORF46.1 的杂交蛋白；(f)含 ORF46.1 和 919 的杂交蛋白；(g)含 ORF46.1、287 和 919 的杂交蛋白；(h)含 919 和 519 的杂交蛋白；和(i)含 ORF97 和 225 的杂交蛋白。图 14 显示了其它实施例。

当使用 287 时，其优先位于杂交体的 C-末端；如果在 N-末端使用它，则
5 优先使用 287 的 Δ G 形式(如与 ORF46.1、919、953 或 961 杂交的杂交体的 N-末端)。

当使用 287 时，其优先为菌株 2996 或菌株 394/98 的。

当使用 961 时，其宜在 N-端。可使用 961 的结构域形式。

WO 99/66741 公开了 ORF46、287、919 和 953 的多聚体形式排序。本发明
10 可以使用这些多聚体形式中的任一形式。

温度

在异源表达的第六个步骤中，低温表达本发明的蛋白质。

表达的奈瑟球菌蛋白(如 919)可能对大肠杆菌有毒性，这可以通过在不显
15 现蛋白质毒性的温度时表达该蛋白来避免。

因此，本发明提供了异源表达本发明蛋白质的方法，其中在不显现该蛋白质毒性的温度时进行本发明蛋白质的表达。

优选的温度为约 30℃。这特别适用于 919 的表达。

20 突变

如上所述，表达的奈瑟球菌蛋白质对大肠杆菌可能有毒性的。通过突变该蛋白以降低或消除毒性来避免这种毒性。特别是，可以使用降低或消除毒性酶活性的突变，较佳地使用定位诱变。

因此在异源表达的第七个步骤中，表达的蛋白质是突变的，以减少或消除
25 毒性。

因此，本发明提供了异源表达本发明蛋白质的方法，其中蛋白质是突变的以降低或消除毒性。

较佳地，该方法适用于蛋白质 907、919 或 922 的表达。在 907 中的优选突变是在 Glu-117(如 Glu→Gly)；在 919 中的优选突变是在 Glu-255(如 Glu→

Gly)和/或在 Glu-323(如 Glu→Gly);在 922 中的优选突变是在 Glu-164(如 Glu→Gly)、Ser-213(如 Ser→Gly)和/或 Asn-348(如 Asn→Gly)。

其它载体

- 5 在异源表达的第八个步骤中,用其它载体表达蛋白质。这可以改进表达收率,例如,或使用已验证用于 GMP 的质粒。

因此,本发明提供了一种异源表达本发明蛋白质的方法,其中使用其它载体。所述的其它载体最好为 pSM214,其无融合配体。可以包含或不包含无前导肽。

- 10 这种方法特别适用于蛋白质 953。用从 pSM214 表达的 953 天然前导肽对 953 的表达和定位比从 pET 载体的好得多。

也可将 pSM214 与以下一起使用:△G287、△2-287,△3-287,△4-287,Orf46.1,961L,961,961(MC58),961c-L,919,953 和△G287-Orf46.1。

- 15 其它适合的载体是 pET-24b(Novagen;用卡那霉素抗性),而且也不使用融合配体。pET-24b 宜与以下一起使用:△G287K,△2-287K,△3-287K,△4-287K,Orf46.1,Orf46A-K,961-K(MC58),961a-K,961b-K,961c-K,961c-L-K,961d-K,△G287-919-K,△G287-Orf46.1-K 和△G287-961-K。

- 20 多聚体形式

在异源表达的第九步中,表达或纯化蛋白质使其接受一种特定的多聚体形式。

这种方法尤其适用于蛋白质 953。953 的一种特定的多聚体形式(单体形式)的纯化使蛋白质比其它形式(二体形式)具有更大的杀菌活性。

- 25 可将蛋白质 287 和 919 以二体的形式纯化。

可将蛋白质 961 以 180kDa 寡聚形式(如四体)纯化。

脂质化

在异源表达的第十步中,将蛋白质以脂质化的蛋白质表达。

因此，本发明提供了一种异源表达本发明蛋白质的方法，其中蛋白质是以脂化的蛋白质表达的。

这尤其适用于 919、287、ORF4、406、576-1 和 ORF25 的表达。WO 00/66741 公开了 919、287 和 ORF4 的多聚体形式。

5 该方法通常包括使用合适的前导肽而不用 N-末端融合配体。

C-末端缺失

在异源表达的第十一步中，使本发明蛋白质的 C-末端突变。另外，不宜使用融合配体。

10 因此，本发明提供了异源表达本发明蛋白质的方法，其中 (a) 使该蛋白质的 C-末端区域突变，并任选地 (b) 不使用融合配体。

这种方法通常包括如下步骤：获得编码本发明蛋白质的核酸；操作所述的核酸从而使编码蛋白质 C-末端部分的核苷酸突变。将得到的核酸插入表达载体中，或作为表达载体的一部分。表达的蛋白质的第一氨基酸为成熟天然蛋白质的第一氨基酸。

15 的第一氨基酸。

这种突变可以是取代、插入或较佳地为缺失。

这种方法能增加表达水平，尤其是蛋白质 730、ORF29 和 ORF46 的表达水平。对蛋白质 730 而言，可以删除约 65-214 氨基酸的 C-末端区域；对 ORF46 而言，可以删除约 175 氨基酸的 C-末端区域；对 ORF29 而言，可以删除 C-末端以留下约 230-370 N-末端氨基酸。

20 端以留下约 230-370 N-末端氨基酸。

前导肽突变

在异源表达的第十二步中，使蛋白质的前导肽突变。这尤其适用于蛋白质 919 的表达。

25 因此，本发明提供了一种异源表达本发明蛋白质的方法，其中使蛋白质前导肽突变。

这种方法通常包括如下步骤：获得编码本发明蛋白质的核酸；并操作所述的核酸使前导肽中的核苷酸突变。可将得到的核酸插入表达载体中，或作为表达载体的一部分。

聚-甘氨酸缺失

在异源表达的第十三步中，使野生型序列的聚-甘氨酸段突变。以此提高蛋白质的表达。

- 5 聚-甘氨酸段包括序列(Gly)_n，其中 $n \geq 4$ (如 5、6、7、8、9 或更多)。让该段突变以破坏或除去(Gly)_n。这可以通过删除(如，CGGGGS→CGGGS、CGGS、CGS 或 CS)，通过替代(如 CGGGGS→CGXGGS、CGXXGS→CGXGXS 等)，和/或缺失(如 CGGGGS→CGGXGGS、CGXGGGS 等)。

- 这种方法不局限于奈瑟球菌蛋白-其可用于任何蛋白质(尤其是细菌蛋白)
- 10 以提高异源表达。然而，对奈瑟球菌蛋白质而言，它特别适用于表达 287、741、983 和 Tbp2。W0 00/66741 公开了 287 的多聚体形式的排序。

因此，本发明提供了一种异源表达本发明蛋白质的方法，其中(a)使蛋白质中的聚-甘氨酸段突变。

- 这种方法通常包括如下步骤：获得编码本发明蛋白质的核酸；和操作所述
- 15 的核酸使编码蛋白质序列中的聚-甘氨酸段的核苷酸突变。可将得到的核酸插入表达载体中，或作为表达载体的一部分。

相反，可用相反的方法(即引入聚-甘氨酸段)抑制或减少特定异源蛋白质的表达。

20 异源宿主

- 虽然本发明蛋白质的表达可以在天然宿主(即天然表达蛋白质的生物)中发生，但本发明用异源宿主。异源宿主可以是原核细胞或真核细胞。最好是大肠杆菌，其它合适的宿主包括枯草杆菌、霍乱弧菌、伤寒沙门氏菌(*Salmonella typhi*)、鼠伤寒沙门氏菌(*Salmonella typhimurium*)、脑膜炎奈瑟球菌、淋病
- 25 奈瑟球菌、乳糖奈瑟球菌(*Neisseria lactamica*)、灰色奈瑟球菌(*Neisseria cinerea*)、分枝杆菌(如结核分枝杆菌(*M. tuberculosis*))、酵母等。

载体等

如上所述的方法，本发明提供了(a)用于这些方法的核酸和载体；(b)含所

- 述载体的宿主细胞；(c)可用这些方法表达的或可表达的蛋白质；(d)包含这些蛋白质的组合物，其可能适用于疫苗、或例如诊断剂或免疫原性组合物；(e)用作药物(如疫苗)或诊断剂的组合物；(f)这些组合物用于生产(1)治疗或预防奈瑟球菌引起的感染的药物(2)检测奈瑟球菌或由奈瑟球菌引起的抗体存在与否的诊断剂，和/或(3)能产生抗奈瑟球菌抗体的药物；和(g)治疗患者的方法，其包括对该患者施用治疗有效量的这些组合物。

序列

- 本发明还提供了具有以下实施例中所列出的任何序列的蛋白质或核酸。本发明还提供了具有与这些序列是序列相同性的蛋白质和核酸。如上所述，“序列相同性”的程度最好大于50%(如60%、70%、80%、90%、95%、99%或更高)。

另外，本发明还提供了能与实施例中公开的核酸杂交的核酸，较佳地在“高度严谨”条件下(如在65℃，0.1×SSC，0.5%SDS溶液中)。

- 15 本发明还提供了编码本发明蛋白质的核酸。

还应理解，本发明提供了核酸，所述核酸含有与上述序列互补的序列(例如用于反义或探测目的)。

- 当然，本发明的核酸可以用多种方法制备(例如，化学合成、从基因组或cDNA文库、或从生物体本身制得等)，并可以是各种形式(例如单链、双链、
20 载体、探针等)。

此外，术语“核酸”包括DNA和RNA，以及它们的类似物，如含有修饰骨架的那些类似物，还包括肽核酸(PNA)等。

附图简述

- 25 图1和2显示用于以异源前导肽表达蛋白质的构建物。

图3显示ORF1的表达数据，图4显示蛋白质961的类似数据。

图5显示蛋白质287的结构域，图6和7显示结构域A中的缺失。

图8显示蛋白质564的结构域。

- 图9显示由919前导肽驱动的PhoC报道基因，图10显示用前导肽的突变
30 体得到的结果。

图 11 显示蛋白质 730 的插入突变体(A:730-C1; B: 730-C2)。

图 12 显示蛋白质 961 的结构域。

图 13 显示 ΔG 蛋白质的 SDS-PAGE。点显示主要重组产物。

图 14 显示本发明的 26 杂交蛋白。

5

本发明的进行模式

实施例 1-919 及其前导肽

脑膜炎奈瑟球菌(血清群 B, 菌株 2996)的蛋白质 919 具有如下序列:

```

1  MKKYLFRAL YGIAAILAA CQKSIQTFP QPDTSVINGP DRPVGIPDPA
51  GTTVGGGGAV YTVVPHLSLP HMAAQDFAKS LQSFRLGCAN LKNRQGWQDV
101 CAQAFQTPVH SFQAKQFFER YTFPWQVAGN GSLAGTVTGY YEPVLKGDDR
151 RTAQARFPIY GIPDDFISVP LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT
201 HTADLSRFPI TARTTAIKGR FEGRFLFYH TRNQINGGAL DGKAPILGYA
251 EDPVELFFMH IQGSGRLKTP SGKYIRIGYA DKNEHPYVSI GRYMADKGYL
301 KLGQTSMQGI KAYMRQNPQR LAEVLGQNPS YIFFRELAGE SMDGPFVGALG
351 TPLMGEYAGA VDRHYITLGA PLFVATAHEFV TRKALNRLIM AQDTGSAIKG
401 AVRVDYFWGY GDEAGELAGK QKTTGYVWQL LPNGMKPEYR P*

```

10 在前导肽下加下划线。

可以在 WO 00/66741 的图 7 和 18 中发现其它菌株的 919 序列。

WO 99/57280 的实施例 2 公开了在大肠杆菌中作为 His-融合体的蛋白质 919 的表达。该蛋白质是优良的外露表面免疫原。

将另三种表达方法用于 919:

15 1)919 没有其前导肽(且没有成熟的 N-末端半胱氨酸)且无任何融合配体 ('919 未标记的):

```

1  QSKSIQTFP QPDTSVINGP DRPVGIPDPA GTTVGGGGAV YTVVPHLSLP
50  HMAAQDFAKS LQSFRLGCAN LKNRQGWQDV CAQAFQTPVH SFQAKQFFER
100 YTFPWQVAGN GSLAGTVTGY YEPVLKGDDR RTAQARFPIY GIPDDFISVP
150 LPAGLRSGKA LVRIRQTGKN SGTIDNTGGT HTADLSRFPI TARTTAIKGR
200 FEGRFLFYH TRNQINGGAL DGKAPILGYA EDPVELFFMH IQGSGRLKTP
250 SGKYIRIGYA DKNEHPYVSI GRYMADKGYL KLGQTSMQGI KAYMRQNPQR
300 LAEVLGQNPS YIFFRELAGE SMDGPFVGALG TPLMGEYAGA VDRHYITLGA
350 PLFVATAHEFV TRKALNRLIM AQDTGSAIKG AVRVDYFWGY GDEAGELAGK
400 QKTTGYVWQL LPNGMKPEYR P*

```

通过设计从预定的前导序列的 5'-端扩增引物下游除去前导肽和半胱氨酸。

2)包含其前导肽但无任何融合配体的 919('919L'); 和

20 3)包含 ORF4 的前导肽(MKTFFKTL SAAALALILAA)的 919('919Orf4')。

```

1  MKTFFKTLG AAALALILAA CQSKSIQTTP QPDTSVINGP DRPVGIPDPA
50  GTTVGGGGAV YTVVPHLSLP HMAAQDFAKS LQSFRLGCAN LKMRQGMQDV
100 CAQAFQTFVH SFQAKQFFER YFTPMQVAGN GSLAGTVTGY YEFVLKGDDR
150 RTAQARFPTY GIPDDFISVF LPAGLRSGKA LVRIRQTGEN SOTIDNTGGT
200 ETADLSRFFI TARTTAIKGR FEGSRFLPYH TRNQINGGAL DGEAPILGYA
250 EDPVELFFMH IQGSORLKTP SGKYIRIGYA DQNEHPYVSI GRYMADKGYL
300 KLGQTSMQGI KSYMQRNPQR LAEVLGQNPS YIFFRELAGE SNDGPVGALG

350 TPLMGEYAGA VDRHYITLGA PLFVATAHFV TRKALNRLIM AQDTGSAIKG
400 AVRVDYFWGY GDEAGELAGK QKTTGYVMQL LPNGMKPEYR P*

```

为了制备此构建物, 编码 ORF4 前导肽的整个序列作为尾部包括于 5'-引物 (引物 919L 或 f4For)。用编码 ORF4 前导序列的序列中的双核苷酸改变产生 NheI 5 限制酶切位点(无氨基酸改变), 使不同的基因融合于 ORF4 前导肽序列。在所有的 3'-端引物序列中包括终止密码子。

蛋白质的所有三种形式都被表达且可以纯化。

如所示的³H]-棕榈酸标记的掺入, 将 '919L'和 '919Orf4'表达产物脂化。919 未标记的不掺入 ³H 标记, 且是在细胞内定位的。

10 与 919L 相比, 919Lorf4 更易纯化。将其纯化并用于免疫接种小鼠。得到的血清在 FACS 和 ELISA 测试以及杀菌试验中产生极佳的结果。该脂蛋白显示位于外膜。

919 未标记的具有极佳的 ELISA 滴定度和高血清杀菌活性。FACS 证明了其细胞表明的定位。

15

实施例 2-919 和表达温度

于 37°C, 表达 919Lorf4 蛋白质的大肠杆菌的生长导致细菌裂解。为了克服这个问题, 重组细菌在 30°C 生长。防止了裂解, 而不会制止表达。

20 实施例 3-907、919 和 922 的突变

假设蛋白质 907、919 和 922 是胞壁质水解酶和具体说是裂解性转糖基酶。胞壁质裂解酶位于外膜, 且参与肽聚糖的降解。

因此测试纯化的蛋白质 919 未标记的、919Lorf4、919-His(即, 包含 C-末端 His-标记)和 922His 的胞壁质水解酶活性[Ursinus & Holtje(1994)J.Bact.176:338-25 343]。使用两种不同的试验, 一种用于确定不溶性胞壁质小囊降解为可溶性胞壁肽的, 另一种则是测定聚(MurNAc-GlcNAc)_{n>30} 聚糖链的断裂。

第一种试验使用由内消旋-2,6-二氨基-3,4,5- ^3H 庚二酸放射标记的胞壁质小囊作为底物。于 37°C ，在总体积为 $100\mu\text{l}$ (包括 10mM Tris-马来酸盐($\text{pH } 5.5$)、 10mM MgCl_2 、 0.2% v/v Triton X-100 和 ^3H A2pm 标记的胞壁质小囊(约 10000cpm))中培养酶(总共 $3\text{--}10\mu\text{g}$)45 分钟。将试验混合物置于冰上 15 分钟，
5 在 $100\mu\text{l}$ 1% w/v N-乙酰基-N,N,N-三甲铵上 15 分钟，从 10000g 离心 15 分钟沉淀物质。用液体闪烁计数测定上清液中的放射活性。用大肠杆菌可溶性裂解性转糖基酶 Slt70 作为试验的阳性对照；阴性对照包括不含酶的上述试验溶液。

在第一个试验中，除 919-His 以外所有蛋白质都产生阳性结果。

第二种试验监测聚(MurNAc-GlcNAc)聚糖链的水解。于 37°C ，在 10mM Tris-马来酸盐($\text{pH } 5.5$)、 10mM MgCl_2 和 0.2% v/v Triton X-100 中，将 N-乙酰基-D-1- ^3H 葡萄糖胺标记的纯化的链聚(MurNAc-GlcNAc) $_{n>30}$ 与 $3\mu\text{g}$ 919L 一起培育。煮沸 5 分钟终止反应，加入 $10\mu\text{l}$ 20% v/v 磷酸将样品的 pH 调节至约 3.5。如 Harz 等 [Anal.Biochem.(1990)190:120-128]所述，在 Nucleosil 300 C_{18} 柱上通过反相 HPLC 将底物与产物分开。在本试验中将大肠杆菌裂解性转糖基酶 Mlt A
15 作为阳性对照。无酶进行的试验作为阴性对照。

当用 HPLC 将无水双糖亚基从寡糖中分开时，用这种试验证明了 919Lorf4 水解分离的聚糖链的能力。

选择蛋白质 919Lorf4 用于动力学分析。在试验缓冲液中添加的 0.2% v/v Triton X-100 使 919Lorf4 的活性提高 3.7 倍。Triton X-100 的存在对 919 未标记的
20 活性没有影响。在范围为 5.0-8.0 的 Tris-马来酸盐缓冲液中测定了 pH 对酶活性的影响。反应的最适 pH 为 5.5。在 18°C 到 42°C 的温度范围中，在 37°C 观察到最大活性。在终浓度为 10mM 的各种离子的存在下进行反应，以测定各种离子对胞壁质水解酶活性的影响。发现存在 Mg^{2+} 时活性最高，其刺激活性 2.1 倍。 Mn^{2+} 和 Ca^{2+} 还刺激酶活性至类似程度，而添加 Ni^{2+} 和 EDTA 没有明显作用。相反， Fe^{2+} 和 Zn^{2+} 显著抑制酶活性。
25

如 Glauner[Anal.Biochem.(1988)172:451-464]所述，用反相 HPLC 分析由消化未标记的大肠杆菌胞壁质小囊得到的反应产物的结构。用由溶菌酶 Cellosyl 消化的胞壁质小囊来校准和标定 Hypersil ODS 柱。主要的反应产物是 1,6-脱水双糖四和三肽，表明 1,6-无水胞壁质酸(muraminic)分子内键的形成。

这些结果从实验角度表明, 919 是胞壁质水解酶, 尤其是裂解性转糖基酶家族的成员。另外, 922-His 裂解胞壁质小囊的能力表明该蛋白质也是裂解性转糖基酶。

这种活性可有助于解释 919 在大肠杆菌内表达时的毒性作用。

- 5 为了消除酶促活性, 使用合理的诱变。907、919 和 922 显示对大肠杆菌的三种膜-结合的、脂化的胞壁质裂解性转糖基酶的相当低的同源性:

919(441aa)与大肠杆菌 MLTA(P46885)为 27.3% 相同, 超过 440aa 重叠;

922(369aa)与大肠杆菌 MLTB(P41052)为 38.7% 相同, 超过 310aa 重叠; 和

907-2(207aa)与大肠杆菌 MLTC(P52066)为 26.8% 相同, 超过 149aa 重叠。

- 10 907-2 还与大肠杆菌 MLTD(P23913)和 Slt70(P03810)具有同源性, 它是一种位于周质间隙中的可溶性裂解性转糖基酶。在 919、922 和 907-2 之间没有测出明显的序列同源性, 且在相应的 MLTA、MLTB 和 MLTC 蛋白质之间的情况也同样如此。

- Slt70([1QTEA; 1QTEB; Thunnissen 等 (1995)Biochemistry 34:12729-12737]
15 和 Slt35[1LTM;1QUS; 1QUT; van Asselt 等 (1999) Structure Fold Des 7:1167-80] (40kDa MLTB 的可溶形式)可有晶体结构。

已鉴定了 Slt70 和 MLTB 的催化残基一种(谷氨酸)。

- 对 Slt70 而言, 诱变研究表明即使用谷氨酰胺保守性地替代催化 Glu505 也会导致酶活性的完全丧失。虽然 Slt35 与 Slt70 没有明显的序列相似性, 但它们的
20 催化区域却显示出令人惊讶的相似性。MLTB 中相应的催化残基是 Glu162。

另一个被认为在酶裂缝的正确折叠中起重要作用的残基是谷氨酸的保守良好的甘氨酸(Gly)下游。最近, Terrak 等 [Mol.Microbiol.(1999)34:350-64]提出存在另一种重要的残基, 它是位于催化性谷氨酸约 70-75 残基下游的芳族氨基酸。

- 25 进行 Slt70 与 907-2 的序列对比和 MLTB 与 922 的序列对比, 以鉴定 MenB 抗原中相应的催化性残基。

催化区域的两个序列对比报道如下:

907-2/Slk70:

```

          90      100      110      ▼120      130      140
907-2.pep  EERRLLVNIQYESSRAG--LDTQIVLGLIEVESAFRQYAISSVQARGLMQVMPFNKNIYIG
          ||  |  |  ::  :  |  :  :  ::  :  |||:  :  |  ||  |||:  ||  ::
slty_ecoli ERFFLAYNDLPKRYTSGKEIPQSYAMAIARQESAMNPKVKSPVQASGLMQIMPOTATHTV
          480      490      500      ▲ 510      520      530
                      GLU505

```

922/MLTB

```

          150      160      ▼ 170      180      190      200
922.pep    VAQKYGVPAKLIVAVIGIETNYGENTQSFVRADALATLGFDPFRAGFFQKELVELLKLAA
          :  |  |||  |  :  ||:  ||:  ||  :  |  :  |  |||  ||:  ||:  ||  :  |  :  |
mltb_ecoli ANQVYGVPPETIIVGIIGVETRNQVRMGKTRILDALATLSFNYPRAEYFSGELSTFLMA
          150      160      ▲ 170      180      190      200
                      GLU162

```

```

          210      220      230      240      250      260
922.pep    KEEGGDVFAFKGSYAGAMQMPQFMPSSYRKMAVDYDGDGHRDINGNVGDVAASVANYMKQ
          ::  |  |  :  :  ||:  |||  ||:  |||  ||:  ||:  ||:  ||:  ||  :  |  :  |||  ||:  |
mltb_ecoli RDEQDDPLNLKGSFAGAMGYGQFMPSSYKQYAVDFSGDGHINLNDPVD-DAIGSVANYFKA
          210      220      230      240      250      260

```

从这些对比得出 907-2 中相应的催化谷氨酸盐是 Glu117, 而 922 中为 Glu164。这两种抗原都有下游甘氨酸(在酶裂缝(粗体)折叠中起结构作用), 922 在约 70aa 下游(粗体)具有保守的芳族残基。

- 5 对蛋白质 919 而言, 对大肠杆菌的同源 MLTA 没有 3D 结构, 且对可能的催化性残基的了解为零。然而, 通过与 MLTA 的序列对比, 预计 919 中的三个氨基酸为催化残基:

919/MLTA

```

          240      250      ▼ 260  □ □ 270 □ 280      290
919.pep    ALDGKAPILGYAEDPFVELFFMHIQSGRLKTPSGKYIRI-GYADKNEHPYVSIGRYMADK
          ||:  |  ||:  ||:  ::  :  |  :  |||  :  :  :  :  :  ||  ||  |  |||:  :  |
mlta_ecoli.p ALSDKY-ILAYSNSILMDNFIMDVQSGYIDFGDGSFLNFTSYAGKNGHAYRSIGKVLIDR
          170      180      190      200      210

          300      310      320      ▼ 330 □ □ 340      350  0
919.pep    GYLKLGQTSMQGIKSYMRQNPQ-RLAEVLGQNPYSYIFFRELAGSSNDGFPV-GALGTPLMG
          |  :  |  :  |||  ||:  :  :  :  :  :  ||  |||  ||:  ||:  :  :  ||  ||  :  ||:  |
mlta_ecoli.p GEVKKEDMSMQAIRHWGETHSEAEVRELLEQNPSFVFFKQSFSA----PVKGASAVPLVG
          220      230      240      250      260      270

          360      ▼ 370  o 380      390      400      00410
919.pep    EYAGAVDRHYITLGAPLFVATAHPVTRKALN-----RLIMAQDTGSAIKGAVRVDFYFWGY
          :  :  |  ||  |  |  :  :  :  :  :  :  :  |  ||:  ||  ||:  |||  :  |  :  |
mlta_ecoli.p RASVASDRSIIPPQTLLAEVPLLDNNGKFNGQYELRLMVALDVGGAIKQGQ-HFDIYQGI
          280      290      300      310      320      330

          420      o
919.pep    GDEAGELAGKQRTTGYVWQLLP
          |  |||  :  ||  :  |  ||  |
mlta_ecoli.p GPEAGHRAGWYNHYGRVWVLKT
          340      350

```

- 10 用符号 ▼ 显示三个可能的催化残基:

1) Glu 225(MLTA 中的 Asp), 其后为三个保守性甘氨酸(Gly263、Gly265

和 Gly272)和三个保守性芳族残基(位于约 75-77 残基下游)。这些下游残基以□显示。

2)Glu323(在 MLTA 中保守的),其后为 2 个保守的甘氨酸(Gly347 和 Gly355)和两个保守的芳族氨基酸(位于 84-85 残基下游)(Tyr406 或 Phe407)。用◇显示

5 这些下游残基。

3)Asp362(替代预计的 Glu),其后为一个甘氨酸(Gly369)和一个保守的芳族残基(Trp428)。用 o 显示这些下游残基。

WO00/66741 中公开了 919 的多聚体形式的排序。

基于对催化残基的预测,已产生了 919 的三种突变体和 907 的一种突变体,

10 各包含单个氨基取代。用基于 PCR 的 SDM,以甘氨酸残基替代 919 蛋白质的 255 位和 323 位的谷氨酸、362 位的天冬氨酸和 907 蛋白质的 117 位的谷氨酸。为了达到该目的,设计了包含从 Glu 或 Asp 变为 Gly 密码子的内部引物:

引物	序列	密码子变化
919-E255 for	CGAAGACCCCGTC <u>Ggt</u> CTTTTTTTATG	GAA → Ggt
919-E255 rev	GTGCATAAAAAAAGacCGACGGGGTCT	
919-E323 for	AACGCCTCGCC <u>Ggt</u> GTTTTGGGTCA	GAA → Ggt
919-E323 rev	TTTGACCCAAAACacCGGCGAGGCG	
919-D362 for	TGCCGGCGCAGTC <u>Ggt</u> CGGCACTACA	GAC → Ggt
919-D362 rev	TAATGTAGTGCCGacCGACTGCGCCG	
907-E117 for	TGATTGAGGTG <u>Ggt</u> AGCGCGTTCCG	GAA → Ggt
907-E117 rev	GGCGGAACGCGCTacCCACCTCAAT	

有下划线的核苷酸密码是甘氨酸的,突变的核苷酸为小写字母。

15 为了产生 919-E255、919-E323 和 919-E362 突变体,用 20ng pET-LOrf4 DNA 作为模板和以下的引物对进行 PCR:

- 1)Orf4L 正向/919-E255 反向
- 2)919-E255 正向/919L 反向
- 3)Orf4L 正向/919-E323 反向
- 20 4)919-E323 正向/919L 反向
- 5)Orf4L 正向/919-D362 反向
- 6)919-D362 正向/919L 反向

用 PCR1-2、3-4 或 5-6 的产物作为模板和正向引物以及反向引物(“Orf4L

正向”和“919L 反向”), 进行第二轮 PCR。

对于突变体 907-E117, 用 200ng 2996 菌株的染色体 DNA 作为模板和以下的引物对进行 PCR:

7)907L 正向/907-E117 反向

5 8)907-E117 正向/907L 反向

用 PCR7 和 8 的产物作为模板和寡“907L 正向”和“907L 反向”作为引物, 进行第二轮 PCR。

用以下的标准方法加工包含各突变的 PCR 片段, 用 NdeI 和 XhoI 限制酶消化并克隆入 pET-21b+载体中。用序列分析验证各突变的存在情况。

10 类似地进行 907 中的 Glu117 向 Gly 的突变, 以及 922 中的残基 Glu164、Ser213 和 Asn348 的突变。

919 的 E255G 的突变体显示活性低降 50%; E323G 突变体显示活性降低 70%; E362G 突变体显示活性没有降低。

15 实施例 4-多聚体形式

将 287-GST、919 未标记的和 953-His 进行凝胶过滤以分析四级结构或制备性目的。用 FPLC Superose 12(H/R 10/30)或 Superdex 75 凝胶过滤柱(Pharmacia)计算天然蛋白质的分子量。用于 287、919 和 953 层析的缓冲液分别是 50mM Tris-HCl(pH 8.0)、20mM Bicine(pH 8.5)和 50mM Bicine(pH 8.0)。

20 另外, 各缓冲液包含 150-200mM NaCl 和 10% v/v 甘油。用适合的缓冲液透析蛋白质并以 200 μ l 体积施用。以 0.5-2.0ml/分钟的流速进行凝胶过滤, 在 280nm 监测洗脱物。收集组分, 用 SDS-PAGE 分析。用蓝色葡聚糖 2000 和分子量标准核糖核酸酶 A、胰凝乳蛋白酶 A 卵清蛋白、白蛋白(Pharmacia)校准柱。由标准的 K_{av} 与 $\log M_r$ 的校准曲线计算样品的分子量。凝胶过滤之前, 用凝血酶消化 287-GST 以切割 GST 组成成分。

25

287、919 和 953-His 的估计分子量分别为 73kDa、47kDa 和 43kDa。这些结果表明 919 是单体而 287 和 953 主要是天然二聚的。对 953-His 而言, 在凝胶过滤的过程中观察到两个高峰。主峰(80%)代表 953 的二聚形式, 而次峰(20%)则为单体预计的大小。发现 953 单体形式比二聚形式具有更大的杀菌活性。

实施例 5-pSM214 和 pET-24b 载体

由 pET 载体和 pSM214 表达了包含天然前导肽而无融合配体的 953 蛋白质 [Velati Bellini 等, (1991)J.Biotechnol.18,177-192]。

- 5 用大肠杆菌 MM294-1 菌株作为宿主, 以全长基因将 953 序列克隆入 pSM214 中。为了实现此目的, 用如下引物通过 PCR 扩增 953 基因的整个 DNA 序列(从 ATG 到终止密码子):

953L for/2	<u>CCGGAATTCTTATGAAAAAATCATCTTCGCCGC</u>	Eco RI
953L rev/2	<u>GCCCAAGCTTTTATTGTTTGGCTGCCTCGATT</u>	Hind III

- 10 其分别包含 EcoRI 和 HindIII 限制酶切位点。用 EcoRI 和 HindIII 消化扩增的片段, 并将其与用这两种酶消化的 pSM214 载体接连。将连接的质粒转化入大肠杆菌 MM294-1 细胞中(通过 37℃, 在冰上培育 65 分钟), 并将细菌细胞涂布于包含 20μg/ml 氯霉素的 LB 琼脂上。

- 15 于 37℃, 在包含 20μg/ml 氯霉素的 4ml LB 肉汤中使重组菌落生长过夜; 将细菌细胞离心, 提取质粒 DNA, 并用 EcoRI 和 HindIII 限制进行分析。为了分析重组集落表达蛋白质的能力, 将它们在包含 20μg/ml 氯霉素的 LB 肉汤中培育, 并使它们在 37℃ 生长 16 小时。离心细菌细胞, 悬浮于 PBS 中。用 SDS-PAGE 和考马斯蓝染色分析蛋白质的表达。

pSM214 质粒的表达水平出人意料的高。

用于将序列克隆入 pSM-214 载体中的寡(核苷酸)如下:

ΔG287 (pSM-214)	Fwd	CCGGAATTCTTATG-TCGCCCAGATGTTAAATCGGCGGA	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Δ2 287 (pSM-214)	Fwd	CCGGAATTCTTATG-AGCCAAGATATGGCGGCAGT	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Δ3 287 (pSM-214)	Fwd	CCGGAATTCTTATG-TCCGCCGAATCCGCAAATCA	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Δ4 287 (pSM-214)	Fwd	CCGGAATTCTTATG-GGAAGGGTTGATTGGCTAATG	EcoRI
	Rev	GCCCAAGCTT-TCAATCCTGCTCTTTTTTGCCG	HindIII
Orf46.1 (pSM-214)	Fwd	CCGGAATTCTTATG-TCAGATTGGCAAACGATTCTT	EcoRI
	Rev	GCCCAAGCTT-TTACGTATCATATTTACGTGCTTC	HindIII
ΔG287-Orf46.1 (pSM-214)	Fwd	CCGGAATTCTTATG-TCGCCCAGATGTTAAATCGGCGGA	EcoRI
	Rev	GCCCAAGCTT-TTACGTATCATATTTACGTGCTTC	HindIII
919 (pSM-214)	Fwd	CCGGAATTCTTATG-CAAAGCAAGAGCATCCAAACCT	EcoRI
	Rev	GCCCAAGCTT-TTACGGCGGTATTCGGGCT	HindIII
961L (pSM-214)	Fwd	CCGGAATTCATATG-AAACACTTTCCATCC	EcoRI
	Rev	GCCCAAGCTT-TTACCACTCGTAATTGAC	HindIII
961 (pSM-214)	Fwd	CCGGAATTCATATG-GCCACAAGCGACGAC	EcoRI
	Rev	GCCCAAGCTT-TTACCACTCGTAATTGAC	HindIII
961c L pSM-214	Fwd	CCGGAATTCTTATG-AAACACTTTCCATCC	EcoRI
	Rev	GCCCAAGCTT-TCAACCCACGTTGTAAGGTTG	HindIII
961c pSM-214	Fwd	CCGGAATTCTTATG-GCCACAAACGACGACG	EcoRI
	Rev	GCCCAAGCTT-TCAACCCACGTTGTAAGGTTG	HindIII
953 (pSM-214)	Fwd	CCGGAATTCTTATG-GCCACCTACAAAGTGGACGA	EcoRI
	Rev	GCCCAAGCTT-TTATTGTTGGCTGCCTCGATT	HindIII

如对 953L 所述，操作、克隆和表达这些序列。

对 pET-24 载体而言，如以下对 pET21 所述，克隆这些序列并在 pET-24 中表达这些蛋白质。pET2 具有与 pET-21 相同的序列，但其具有卡那霉素抗性盒替代氨苄青霉素盒。

用于将序列克隆入 pET-24b 中的寡核苷酸是：

ΔG 287 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTAAATCGGC [§]	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGGCC *	XhoI
Δ2 287 K	Fwd	CGCGGATCCGCTAGC-CAAGATATGGCGGCAGT [§]	NheI
Δ3 287 K	Fwd	CGCGGATCCGCTAGC-GCCGAATCCGCAAATCA [§]	NheI
Δ4 287 K	Fwd	CGCGCTAGC-GGAAGGGTTGATTTGGCTAATGG [§]	NheI
Orf46.1 K	Fwd	GGGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCGCTCGAG-TTACGTATCATATTCACGTGC	XhoI
Orf46A K	Fwd	GGGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCGCTCGAG-TTATTCTATGCCTTGTGCGGCAT	XhoI
961 K (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACGA	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961a K	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TCATTTAGCAATATTATCTTTGTTC	XhoI
961b K	Fwd	CGCGGATCCCATATG-AAAGCAAACAGTGCCGAC	NdeI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI
961c K	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961cL K	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCGCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961d K	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	NdeI
	Rev	CCCGCTCGAG-TCAGTCTGACACTGTTTTATCC	XhoI
ΔG 287-919 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCCG	XhoI
ΔG 287-Orf46.1 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TTACGTATCATATTCACGTGC	XhoI
ΔG 287-961 K	Fwd	CGCGGATCCGCTAGC-CCCGATGTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TTACCACTCGTAATTGAC	XhoI

*将该引物用作所有 287 形式的反向引物。

§ 与 ΔG278 K 反向引物联用的正向引物。

5 实施例 6-ORF1 及其前导肽

预计脑膜炎奈瑟球菌(血清群 B, 菌株 MC58)的 ORF1 是用作外膜或分泌的蛋白质。其具有如下序列:

1 MKTDKRTTE THKAPKTGR IRFSPAYLAI CLSFGILPOA WAGHTYFGIN

```

51  YQYYRDFAEK  KGKFAVGAKD  IEVYNKKGEL  VGKSMTKAPM  IDFSVSVSRNG
101  VAALVGDQYI  VSVAHNGGYN  NVDFGABGRN  PDQHRFTYKI  VKRNNYKAGT
151  KGHFYGGDYH  MPRLHKFVTD  AEFVEMTSYM  DGRKYIDQNN  YPDRVRIGAG
201  RQYWRSEDEE  PNNRESSYHI  ASAYSMLVGG  NTFAQNGSGG  GTVNLGSEKI
251  KHSPYGFLLP  GGSFGDSGSP  MFIYDAQKQK  WLINGVLQTG  NPYIGKSNQF
301  QLVKIDWFYD  EIFAGDTHSV  FYEPRQNGKY  SFNDDNNGTG  KINAKHEHNS
351  LPNRLKTRTV  QLFNVSLSET  AREPVYHAAG  GVNSTRPRLN  NGENISFIDE
401  KGKGLILTSN  INQAGGLYF  QGDFTVSPEN  NETWQAGVH  ISEDSTVTNK
451  VNGVANDRLS  KIGKGTLEHV  AKGENQGSIS  VGDGTVILDQ  QADDKGGKQA
501  FSEIGLVSGR  GTVQLNADNQ  FNPDKLYFGF  RGGRLDLNGH  SLSPHRIQNT
551  DEGAMIVNHN  QDKESTVTIT  GNKDIATTGN  NNSLDSKKKI  AYNGMFGKED
601  TTKINGRLNL  VYQPAEDRT  LLLSGGTNLN  GNITQTNGKL  FFSGRPTFHA
651  YNHLNDHWSQ  KEGIPRGEIV  WDNDWINRTF  KAENFQIKGG  QAVVSRNVAK
701  VKGDWHLNHE  AQAVFGVAPH  QSHICTRSD  WTGLTNCVEK  TITDEKVIAS
751  LTKTDISGNV  DLADHAHLNL  TGLATLNGNL  SANGDTRYTV  SHNATQNGNL
801  SLVGNAQATF  NQATLNGNTS  ASGNA SFNLS  DHA VQNGSLT  LSGNAKANVS
851  HSALNGNVSL  ADKAVFHFES  SRFTGQISGG  KDTALHLKDS  EWTLPSTTEL
901  GNINLNDNATI  TLNSAYRHDA  AGAQTGSATD  APRRRSRRSR  RSLSVTPPT
951  SVESRFNTLT  VNGKLNQGT  FRFMSLFGY  RSDKLKLAES  SEGTYTLAVN
1001  NTGNEPASLE  QLTVVEGKDN  KPLSENLFNT  LQNEHVDAGA  WRYQLIRKDG
1051  EFRLHNPVKE  QELSDKLGA  EAKKQAEKDN  AQSLDALIAA  GRDAVKTES
1101  VAEPARQAGG  ENVGIMQAEK  EKKKRVQADK  TALAKQREAE  TRPATTAFFR
1151  ARRARDLPQ  LQPQPQPQP  RDLISRYANS  GLSEFSATLN  SVFAVQDELD
1201  RVFAEDRRNA  VWTSGIRDTK  HYRSQDFRAY  RQQTDLRQIG  MQNHLGSGRV
1251  GILFSEHNTE  NTFDDGIGNS  ARLAGAVFG  QYQIDRFYIG  ISAGAGFSSG
1301  SLSDGIGGKI  RRRVLHYGIQ  ARYRAGFGGF  GIEPHIGATR  YFVQKADYRY
1351  ENVNIATPGL  AFNRYRAGIK  ADYSFKPAQH  ISITPYLSLS  YTDAAAGKVR
1401  TRVNTAVLAQ  DFGKTRSAEW  GVNABIKGFT  LSLHAAAAG  PQLEAQSAG
1451  IKLGYN*

```

前导肽是加下划线的。

WO99/55873 公开了 ORF1 的多聚体形式。

已将三种表达策略用于 ORF1:

- 1) 使用 His 标记的 ORF1, 按照 WO99/24578(ORF1-His);
- 2) 包含自身前导肽但无任何融合配体的 ORF1('ORF1L'); 和
- 3) 包含大肠杆菌 OmpA 的前导肽(MKKTAIAIAVALAGFATVAQA)的 ORF1('Orf1LOmpA'):

```

MKTATAIAVALAGFATVAQAASAGHTYFGINQYYRDFAEKKGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSV
VSRNGVAALVGDQYIVSVAHNGGYNVDFGABGRNPDQHRFTYKIVKNNYKAGTQGHFYGGDYHMPRLHKFVTD
AEFVEMTSYMDGRKYIDQNNYPDRVRIGAGRQYWRSEDEEPNNRESSYHIASAYSMLVGGNTFAQNGSGGGT
VNLGSEKIKHSPYGFLLPTGGSGDSGSPMFIYDAQKQKWLINGVLQTGWNPYIGKSNQFQLVRKIDWFYD
EIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHAAG
GVNSTRPRLNNGENISFIDEKGKGLILTSNINQAGGLYFQGDFTVSPENNETWQAGVHISEDSTVTNK
VNGVANDRLSKIGKGTLEHVAKGENQGSISVGDGTVILDQQADDKGGKQAFSEIGLVSGRGT
VQLNADNQFNPDKLYFGFRGGRLDLNGHSLSPHRIQNTDEGAMIVNHNQDKESTVTITGNKDIATTGN
NNSLDSKKKIAYNGMFGKEDTTKINGRLNLVYQPAEDRTLLLGGTNGNLGNITQTNGKLFSSGRPTFHA
YNHLNDHWSQKEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNHEAQAVFG
VAPHQSHICTRSDWTGLTNCVEKTITDEKVIASLTKTDISGNVLDADHAHLNLTGLATLNGNLSANGDTRY
TVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNA SFNLS  DHA VQNGSLT  LSGNAKANVS
HSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSWTLPSTTELGNINLNDNATITLNSAYRHDA
AGAQTGSATDAPRRRSRRSRSLSVTPPTSVESRFNTLTVNGKLNQGTFRFMSLFGYRSDKLKLAES
SEGTYTLAVNNTGNEPASLEQLTVVVEGKDNKPLSENLFNTLQNEHVDAGAWRYQLIRKDG
EFRLHNPVKEQELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAVKTESVAEPARQAGGENVGIMQAEK
EKKKRVQADKTALAKQREAE TRPATTAFFRARRARDLPQLQPQPQPQQRDLISRYANSGLSEFSATLN
SVFAVQDELDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQNHLGSGRGILFSEHNTE
NTFDDGIGNSARLAGAVFGQYQIDRFYIGISAGAGFSSGSLSDGIGGKI RRRVLHYGIQARYRAGFG
GIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVR
TRVNTAVLAQDFGKTRSAEWGVNABIKGFTLSLHAAAAGPQLEAQSAGIKLGYN*

```

-) 为了制备此构建物, 用 *NheI* 和 *XhoI* 限制酶消化克隆 pET911LOmpA(见如

下), 并纯化相应于携带 OmpA 前导序列的载体的片段(pETLOmpA)。用寡核苷酸 ORF1-正向和 ORF1-反向(分别包括 *NheI* 和 *XhoI* 限制性酶切位点)扩增编码成熟蛋白质的 ORF1 基因, 用 *NheI* 和 *XhoI* 消化并连接到纯化的 pETOmpA 片段(参见图 1)。由 *NheI* 位点引入另一 AS 二肽。

- 5 该蛋白质的所有三种形式都被表达。可纯化 His-标记的蛋白质, 并确证其为表面外露的且可能分泌的(参见图 3)。用这种蛋白质免疫接种小鼠, 得到的血清在杀菌试验中有极佳的结果。

10 将 ORF1LOmpA 作为完整膜纯化, 并固定于内膜和外膜。出人意料地, 针对 ORF1LOmpA 产生的血清比那些针对 His-标记的蛋白质产生的血清显示更好的 ELISA 和抗杀菌特性。

在其定位的地方将 ORF1I 作为外膜纯化。

实施例 7-蛋白质 911 及其前导肽

脑膜炎奈瑟球菌(血清群 B, 菌株 MC58)的蛋白质 911 具有如下序列:

15 1 MKKNILEFNV GLFVLIGAAA VAFLAFRVAG GAAFGGSDKT YAVYADFQDI
 51 GGLKVNAPVK SAGVLVGEVG ATGLDPKSYQ ARVRLDLGK YQFSSDVSAQ
 101 ILTSGLLGEQ YIGLQQGGDT ENLAAGDTIS VTSSAMVLEN LIGKFMTSFA
 151 EKNADGGNAE KAAE*

前导肽是加下划线的。

将三种表达策略用于 911:

- 1) 包含其自身前导肽但无任何融合配体的 911('911L');
 2) 包含大肠杆菌 OmpA 的前导肽的 911('911LOmpA').
 20 为了制备该构建物, 在 5'-引物中包括编码 OmpA 前导肽的完整序列作为尾(引物 911OmpA 正向)。在编码 OmpA 前导肽的序列和编码预计的成熟蛋白质的 911 基因之间插入 *NheI* 限制酶切位点(插入一个氨基酸, 一个丝氨酸), 使该构建物用于克隆不同的基因下游 OmpA 前导肽序列。

- 3) 包含胡萝卜软腐欧文氏菌(*Erwinia carotovora*)PelB 的前导肽
 25 (MKYLLPTAAAGLLLLAAQPAMA)的 911('911LpelB)。

为了制备该构建物, 将 5'-端 PCR 引物设计为前导序列的下游, 并包括 *NcoI* 限制酶切位点, 从而使 911 直接融合于 PelB 前导序列; 3'-端引物包括终止密码子。所用的表达载体是 pET22b+(Novagen), 其携带编码 PelB 前导肽的序列。

NcoI 位点在 PelB 序列后引入另一甲硫氨酸。

表达该蛋白质的所有三种形式。用 911L 的 ELISA 滴定度最高, 919LOmpA 也有良好的结果。

5 实施例 8-ORF46

脑膜炎奈瑟球菌(血清群 B, 菌株 2996)的完整 ORF46 蛋白质具有如下序列:

```

1  LGISRKISLI LSILAVCLPM HAAASDLAND SFIRQVLDRO HFEPDGKYL
51 FGSRGELAER SGHIGLQNTIQ SHQLGNLMIQ QAAIKGNIGY IVRFSDEGHE
101 VHSPFDNHAS HSDSDRAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGY
151 PAPKGARDIY SYDIKGVACN IRLNLTDRS TQQLADRFH HAGSMLTQGV
201 GDGFKRATRY SPFLDRSGNA AEAFNGTADI VKNIIGAAGE IVGAGDAVQG
251 ISEGSENIAM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDMAVQNP
301 NAAQGIEAVS NIFMAAIPK GIGAVRGKYG LGGITAHPIK RSQMGAIALP
351 KGKSAYSINF ADAAYAKYPS PIHSRNIRSM LEQKYGKKN TSSTVPPSNG
401 KNVKLADQRH PKTGVPFDGK GFFNPFKEVK YDTKLDIQEL SGGGIPKAKP
451 VSDAKPRMEV DRKLNKLTTR EQVEKNVQEI RNGMKNSENF QHAQLEREIN
501 KLKSADENIF ADGMGKPTDS MNDKAFSRLV KSVKENGFTN PVVEYVEING
551 KAYIVRGNNR VFAAEYLGR HELEKFKVD PVPNTSWKNP TDVLNESGNV
601 KRPRYRSK*

```

前导肽是加下划线的。

在 WO00/66741 中发现有其它菌株的 ORF46 序列。

10 已将三种表达策略用于 ORF46:

1) 包含自身前导肽但无任何融合配体的 ORF46('ORF46-2L');

2) 不含自身前导肽且无任何融合配体的 ORF46('ORF46-2'), 通过从预计的前导序列设计 5'-端扩增引物下游除去前导肽:

```

1  SDLANDSFIR QVLDQRHFEF DGKYHLPGR GELAERSGHI GLGKIQSHQL
51 GNLMIQQAAL KGNIGYIVRF SDHGHEVHSP FINEASHSDS DRAGSPVDGF
101 SLYRIHWDGY EHHPADGYDG PQGGYPAPK GARDIYSYDI RGVAQNIRLN
151 LTDNRSTGQR LADRFHAGS MLTQGVGDGF KRATRYSPFL DRSGNAAEAF
201 NGTADIVKNI IGAAGEIVGA GDAVQGISEG SNLAVMHGLG LLSTENKMAR
251 INDLADMAQL KDYAAAAIRD NAVQNPNAAQ GIEAVSNIFM AAPIKIGIGA
301 VRGKYGLGGI TAHPIKRSQM GAIALPRGKS AVSONFADAA YAKYPSFYHS
351 RNIRSNLEQR YGKENTTSST VPPSNGKNV LADQRHPKTG VFPDGGKFPN
401 FEKHVKYDTK LDIQELSGGG IPKAKFVSDA KPRMEVDKUL NKLTTRQVE
451 KNVQHIRNGN KNSNFSQHAQ LEREINKLKS ADEINFADGM GEFTDSMDK
501 AFSRLVKSVK ENGFTNFVVE YVEINGKAYI VRGNRVFAA EYLGRHELEK
551 FKKVDFFVPPN TSWKNPTDVL NESQNVKRPR YRSK*

```

15 3) 作为一种截短的蛋白质的 ORF46, 由前 433 氨基酸构成('ORF46.1L'), 通过设计 PCR 引物以扩增相应于 aa1-433 的部分序列构建。

在 3'-端引物序列中包括终止密码子。

对大肠杆菌, ORF46-2L 以很低的水平表达。除去它的前导肽(ORF46-2) 也不能解决这个问题。然而, 截短的 ORF46.1L 形式(前 433 个氨基酸, 在血清

群和品种中很好保守的)能很好地表达,且在 ELISA 和杀菌试验中得到极佳的结果。

也将 ORF46.1 用作杂交蛋白的基础。将其与 287、919 和 ORF1 融合。通常杂交蛋白是不溶的,但能得到一些良好的 ELISA 和杀菌(针对同源 2996 菌株)

5 结果。

蛋白质	ELISA	杀菌 Ab
Orf1-Orf46.1-His	850	256
919-Orf46.1-His	12900	512
919-287-Orf46-His	n.d.	n.d.
Orf46.1-287His	150	8192
Orf46.1-919His	2800	2048
Orf46.1-287-919His	3200	16384

为了比较,构建 ORF46.1、287(作为 GST 融合或以 Δ G287 形式)的‘三’杂交体,针对各种菌株(包括同源 2996 菌株)与三种抗原的简单混合物测试。FCA 用作佐剂:

	2996	BZ232	MC58	NGH38	F6124	BZ133
混合物	8192	256	512	1024	>2048	>2048
ORF46.1-287-919his	16384	256	4096	8192	8192	8192
Δ G287-919-ORF46.1his	8192	64	4096	8192	8192	16384
Δ G287-ORF46.1-919his	4096	128	256	8192	512	1024

10

另外,杂交体显示相等的或更佳的免疫活性。

针对各种异源菌株,将两种蛋白质(菌株 2996)的杂交体与单种蛋白质进行比较:

	1000	MC58	F6124 (MenA)
ORF46.1-His	<4	4096	<4
ORF1-His	8	256	128
ORF1—ORF46.1-His	1024	512	1024

15

杂交体再一次显示相等的或更佳的免疫活性。

实施例 9-蛋白质 961

脑膜炎奈瑟球菌(血清群 B, 菌株 MC58)的完整 961 蛋白质具有如下序列:

```

1  MSMKHFFPAKV LTTAILATFC SGALAATSDD DVKKAATVAI VAAVNNQGEI
51 NGFKAGETTY DIGEDGTTITQ KDATAADV EA DDFKGLGLKK VVTNLTKTVN
101 ENKQNVDAKV KAARSEIEKL TTKLADTDAA LADTDAALDE TTNALNKLGE
151 NITTFARETK TNIVKIDIEL EAVADTVDKH AEAFEDDIADS LDETNTKADE
201 AVKTANEAKQ TARETKQNV DAKVKAETAA GKAEAAAGTA NTADKAEAV
251 AAKVTDIKAD IATNKADIAK NSARIDSLDK NVANLRKETR QGLAEQAALS
301 GLFQFYNVGR FNVTAAVGGY KSESAVAIGT GFRFTENFAA KAGVAUGTSS
351 GSSAAYHGV NYEW*

```

前导肽是加下划线的。

对 961 的表达采用如下三种方法:

- 5 1)使用 GST 融合体的 961, 按照 WO99/57280('GST961');
- 2)包含其自身前导肽但无融合配体的 961('961L'); 和
- 3)无自身前导肽且不含任何融合配体的 961('961 未标记的'), 通过设计从预计的前导序列的 5'-端 PCR 引物下游除去前导肽。

10 该蛋白质的所有三种形式都被表达。可以纯化 GST-融合蛋白, 其抗体表明 961 是表面外露的(图 4)。用该蛋白质免疫接种小鼠, 得到的血清在杀菌试验中有极佳的结果。也可纯化 961L, 它产生非常高的 ELISA 滴定度。

蛋白质 961 似乎是为相变量。另外, 在脑膜炎奈瑟球菌的所有菌株中都没有它。

15 实施例 10-蛋白质 287

脑膜炎奈瑟球菌(血清群 B, 菌株 2996)的蛋白质 287 具有如下序列:

```

1  MFERSVIAMA CIFALSACGG GGGGSPDVKS ADTLSPAP VVAEKETEVI
51 EDAPQAGSQ QGAPSTQGSQ DMAAVSAENT GNGGAATTDK PKNEDEGPQN
101 DMPQNSAESA NQTGNQPAD SSDSAPASNP APANGGSGNFG EVDLANGVLI
151 DGPSQNTITL HCKGDSNGD NLLDEEAPSK SEFENLNESE RIEKYKKGDK

201 SDKFTNLVAT AVQANGTKY VLIYKDKSAS SSSARFRSA RSRRLPAEM
251 PLIPVNQADT LIVDGEAVSL TGHSGNIFAP EGNRYXLYTG AEKLPGGSYA
301 LRVQGEPAKG EMLAGTAVYN GEVLHFHTEN GRPYPTRGRF AAKVDFGSKS
351 VDGIIDSGDD LHMGTQKFKA AIDGNGFKGT WTENGSGDVS GRFYGPAGEE
401 VAGKYSYRPT DAEKGGFGVF AGKKEQD*

```

前导肽是加下划线显示的。

20 在 WO00/66741 的图 5 和 15 中可以发现其它菌株的 287 序列。

WO99/57280 的实施例 9 公开了在大肠杆菌中作为 GST-融合蛋白的 287 的表达。

已使用许多其它方法在大肠杆菌中表达 287，包括：

- 1) 作为 His-标记的融合蛋白的 287('287-His');
- 2) 包含其自身前导肽但无任何融合配体的 287('287L');
- 3) 包含 ORF4 前导肽且无任何融合配体的 287('287LOrf4'); 和
- 4) 不含其自身前导肽且无任何融合配体的 287('287 未标记的');

```

1  CGGGGGGSPD VKSADTLSP AAPVVAEKFT EVKEDAPQAG SQGGGAPSTQ
51  GSQDMAAVSA ENTGNGGAAT TDKPKNEDEG PQNDMPQNSA ESANQTGNNQ
101 PADSSDSAPA SNPAPANGGS NFGRVDLANG VLIDGPSQMI TLTHCKGDSG
151 NGDMILLDEEA PSKSEFENLN ESERIEKYKK DGEKDKFTNL VATAVQANGT
201 NKYVLIYKDK SASSSARFR RSARSRRSLP AEMPLIPVNG ADTLIVDGEA
251 VSLTGHSGNI FAPBGNYRYL TYGAELPGG SYALEVQGEF ANGEMLAGTA
301 VYNGEVLHFF TENGRFYPTR GRFAAKVDFG SKSVDGIIDS GDDLEHGTQK
351 FKAALDGNLF KGTWTENGGS DVSGRFYGPB GSEVAGKYSY RPTDAEKGGF
401 GVFAKKKEQD *

```

所有这些蛋白质都被表达和纯化。

证明 '287L'和 '287LOrf4'是脂蛋白。

如图 2 所示，通过用 *NheI* 和 *XhoI* 消化 919Lorf4 构建 '287Lorf4'。在 5'-
 10 端引物(287Lorf 正向)(融合于 287 编码序列)添加编码遗漏氨基酸的 DNA 序列
 (作为尾)，恢复完整的 ORF4 前导肽。用寡核苷酸 287Lorf4 正向和反向(分别包
 含 *NheI* 和 *XhoI* 位点)扩增编码成熟蛋白质的 287 基因，用 *NheI* 和 *XhoI* 消化并
 连接于纯化的 pETOrf4 片段。

15 实施例 11-包含/不含天然前导肽的其它非融合蛋白质

一种类似的方法适用于大肠杆菌表达 WO99/24578、WO99/36544 和
 WO99/57280 的其它蛋白质。

以下是未用融合配体表达的：008、105、117-1、121-1、122-1、128-1、148、
 216、243、308、593、652、726、982 和 Orf143-1。FACS 表明蛋白质 117-1 是
 20 表面外露的，且产生高 ELISA 滴定度。

以下是由天然前导肽但无融合配体表达的：111、149、206、225-1、235、
 247-1、274、283、286、292、401、406、502-1、503、519-1、525-1、552、556、
 557、570、576-1、580、583、664、759、907、913、920-1、926、936-1、953、
 961、983、989、Orf4、Orf7-1、Orf9-1、Orf23、Orf25、Orf37、Orf38、Orf40、
 25 Orf40.1、Orf40.2、Orf72-1、Orf76-1、Orf85-2、Orf91、Orf97-1、Orf119、Orf143.1。
 这些蛋白质都有后缀 'L'。

用或不用其前导肽表达 His 标记的蛋白质 760。信号肽的缺失大大地增加了表达水平。使用 2M 尿素(用于增溶)最易纯化该蛋白质。

用其自身的信号肽能很好地表达 His-标记的蛋白质 264, 且 30kDa 蛋白质得到阳性 Western 印迹结果。

5 所有蛋白质都成功地表达。

验证了 593、121-1、128-1、593、726 和 982 在细胞质中的定位。

验证了 920-1L、953L、ORF9-1L、ORF85-2L、ORF97-1L、570L、580L 和 664L 在周质中的定位。

验证了 ORF40L 在外膜中的定位, 和 008 和 519-1L 在内膜中的定位。验证了 ORF25L、ORF4L、406L、576-1L 都是在膜中定位的。

发现蛋白质 206 不是脂蛋白。

用其天然前导肽但无融合配体表达的 ORF25 和 ORF40, 以及未用其天然前导肽和融合配体表达的蛋白质 593, 都产生良好的抗杀菌血清。令人惊讶地, 未用融合配体而使用其前导肽(即 'ORF25L'和 'ORF40L')表达的 ORF25 和 ORF40 形式比融合蛋白在杀菌试验中产生更好的结果。

将蛋白质 920L 和 953L 进行 N-末端测序, 结果分别为 HRVWVETAH 和 ATYKVDEYHANARFAF。此测序结果表明切除了预计的前导肽, 且当与周质定位联合时表明: 当由它们的天然前导肽表达时, 大肠杆菌正确地加工和定位这些蛋白质。

20 定位于内膜的蛋白质 519.1L 的 N-末端序列是 MEFFIILLA, 表明未切除前导序列。因此, 它可作为未切割的前导序列和跨膜锚(以与淋病奈瑟球菌的 PBP1 的前导肽类似的方式)[Ropp & Nicholas(1997)J.Bact.179:2783-2787.]。事实上, N-末端区域表现出强疏水性, 且由 Tmpred.程序预计为跨膜的。

25 实施例 12-脂蛋白

用 Kraft 等 [J.Bact.(1998)180:3441-3447.]的方法验证重组脂蛋白中棕榈酸盐的掺入。于 37℃, 在 20ml LB/Amp(100μg/ml)的液体培养基中使携带感兴趣质粒的单集落生长过夜。在 5.0ml 含有 5μC/ml[³H]棕榈酸盐(Amersham)的新鲜培养基 LB/Amp 培养基中, 将培养物稀释至 OD₅₅₀ 为 0.1。当培养物的 OD₅₅₀ 达

- 到 0.4-0.8 时, 用 IPTG(终浓度为 1.0mM)诱导重组脂蛋白 1 小时。用台式离心机以 2700g 离心 15 分钟收获细菌, 用 1.0ml 冷 PBS 洗涤 2 次。将细胞悬浮于 120 μ l 20mM Tris-HCl(pH 8.0)、1mM EDTA、1.0% w/v SDS 中, 并煮沸 10 分钟裂解。以 13000g 离心 10 分钟后, 收集上清液, 加入 1.2ml 冷丙酮沉淀蛋白质, 在 -20 $^{\circ}$ C 静置 1 小时。以 13000g 离心 10 分钟沉淀蛋白质, 重悬浮于 20-50 μ l(以培养物终 O.D 值校准负载量计算的)1.0% w/v SDS。将 15 μ l 等份与 5 μ l SDS-PAGE 样品缓冲液煮沸, 用 SDS-PAGE 分析。电泳凝胶在 10% v/v 乙酸中固定 1 小时后, 在扩增溶液(Amersham)中浸泡 30 分钟。加热真空干燥凝胶, -80 $^{\circ}$ C 暴露在 Hyperfilm(Kodak)中过夜。
- 10 在以下蛋白质中发现 [^{3} H] 棕榈酸盐标记的掺入(表明脂化): Orf4L、Orf25L、287L、287L Orf4、406L、576L、15L、919L 和 919L Orf4。

实施例 13-287 中的结构域

- 根据 287 中不同的结构域与隶属不同功能类别的蛋白质的同源性, 如图 5 所示将其分为 3 个“结构域”。第二个结构显示与 IgA 蛋白酶同源, 第三结构域显示与结合转铁蛋白的蛋白质同源。

- 在脑膜炎奈瑟球菌之间这三个“结构域”各表现出不同程度的序列保守性-结构域 C 为 98% 相同, 结构域 A 为 83% 相同, 而结构域 B 仅为 71% 相同。注意到菌株 MC58 中的蛋白质 287 比菌株 2996 的长 61 个氨基酸。图 7 显示了这两个序列的序列对比, WO00/66741(图 5 和 15)公开了各种菌株的序列对比。

将这三个结构域分别表达成 C-末端 His-标记的蛋白质。用以下构建物对 MC58 和 2996 菌株进行处理:

287a-MC58(aa1-202), 287b-MC58(aa203-288), 287c-MC58(aa311-488)。

287a-2996(aa1-139), 287b-2996(aa140-225), 287c-2996(aa250-427)。

- 25 为了制备这些构建物, 除去 3'-端引物序列中的终止密码子序列。5'引物包括 NheI 限制酶切位点, 3'引物包括作为尾部的 XhoI, 从而用 NdeI-XhoI、NheI-XhoI 或 NdeI-HindIII 限制酶切位点指导将各扩增片段克隆入表达载体 pET21b+ 中。

所有六种构建物都可以表达, 但 287b-MC8 需要变性和重折叠以增溶。

以下描述了结构域 A 的缺失('△4 287-His')。

用菌株 2996 的各种结构域还获得了针对同源和异源 MenB 菌株以及 MenA(F6124 菌株)和 MenC(BZ133 菌株)的免疫学数据(血清杀菌试验):

	2996	BZ232	MC58	NGH38	394/98	MenA	MenC
287-His	32000	16	4096	4096	512	8000	16000
287(B)-His	256	-	-	-	-	16	-
287(C)-His	256	-	32	512	32	2048	>2048
287(B-C)-His	64000	128	4096	64000	1024	64000	32000

5 用菌株 MC58 的结构域获得以下结果:

	MC58	2996	BZ232	NGH38	394/98	MenA	MenC
287-His	4096	32000	16	4096	512	8000	16000
287(B)-His	128	128	-	-	-	-	128
287(C)-His	-	16	-	1024	-	512	-
287(B-C)-His	16000	64000	128	64000	512	64000	>8000

实施例 14-287 中的缺失

与表达单一结构域一样,通过在第一结构域中的渐进缺失还表达了 287(以
10 C-末端 His-标记的蛋白质)。使用菌株 2996 的蛋白质 287 的四缺失突变体(图 6):

- 1) '287-His', 由氨基酸 18-427 构成(即缺失前导肽);
- 2) '△1 287-His', 由氨基酸 26-427 构成;
- 3) '△2 287-His', 由氨基酸 70-427 构成;
- 4) '△3 287-His', 由氨基酸 107-427 构成; 和
- 15 5) '△4 287-His', 由氨基酸 140-427 构成(=287-bc)。

还为菌株 MC58 制备了 '△4'蛋白质('△4 287MC58-His', aa203-488)。

如上所述,用与 287a/b/c 相同的方法制备这些构建物。

可以表达所有六种构建物,并可纯化蛋白质。然而 287-His 的表达很差。

当除去 C-末端 His-标记物时,表达也很高。

20 用缺失突变体也获得了针对同源(2996)和异源 MenB 菌株以及 MenA(F6124 菌株)和 MenC(BZ133 菌株)的免疫学数据(血清杀菌试验):

	2996	BZ232	MC58	NGH38	394/98	MenA	MenC
287-his	32000	16	4096	4096	512	8000	16000
Δ1 287-His	16000	128	4096	4096	1024	8000	16000
Δ2 287-His	16000	128	4096	>2048	512	16000	>8000
Δ3 287-His	16000	128	4096	>2048	512	16000	>8000
Δ4 287-His	64000	128	4096	64000	1024	64000	32000

用菌株 MC58 的序列观察到Δ4 缺失的相同高的活性。同样显示极佳的表达特性，因此突变体是免疫学相等的或更佳的。

5 实施例 15-聚甘氨酸缺失

前实施例的 'Δ1 287-His' 构建物与 287-His 和 '287 未标记的' 的不同仅在于短的 N-末端缺失(GGGGGGS)。然而使用一种以 Nhe 克隆位点中存在的密码子替代缺失的丝氨酸的表达载体，缺失仅为(Gly)₆。因此，该(Gly)₆ 序列的缺失显示对蛋白质表达有明显影响。

- 10 将缺失多达 GGGGGG 的 N-末端氨基酸的蛋白质称为 'ΔG 287'。在菌株 MC58 中，其序列(前导序列加下划线)为：

⇨ ΔG287

```

1  MFKRSVIAMA CIFALSACGG GGGGSPDVKS ADTLSPKPAAP VVSEKETEAR
51  EDAPQAGSQG QGAPSAQGSQ DDAVSEKENT GNGGAVTADN PKHEDVACN
101 DMPQNAAGTD SSTPNETPDP NMLAGMERNQ ATDAGESSQP ANQPTMANAA
151 DGMQGGDPSA GGMNAGNTAA QGANQAGNNQ AAGSSDPFPA SNFAPANGGS
201 NFGKVDLANG VLIDGPSQNI TLTHCKGDSK SGMNPLDEEV QLKSEFEKLS
251 DADKISNYKK DGMNCKFVGL VALSVQMKGI NQYILFYKPK PTSTFARFRS
301 ARSRSLPAE MPLIFVNOAD TLIVDGEAVS LTGESGNIFA PEGNYKYLTY
351 GAEKLPQGSY ALRVQGEPAK GEMLAGAAYV NGEVLHFETE NGRFYPTFRGR
401 FFAKVDGSK SVDGIIDSGD DLEMGTOQFK AALDGNQFGK TWTENGSGDV
451 SGKPYGPAGE EVAGKYSYRP TDAKGGFGV FAGKKEQD*
```

与 '287-His' 或 '287 未标记的' 相比，有或无 His-标记的 ΔG287(分别为 'Δ287-His' 和 'ΔG287K') 以非常好的水平表达。

- 15 以基因变异性数据为基础，许多 MenB 菌株(尤其是菌株 2996、MC58、1000 和 BZ232)的 ΔG287-His 的变体在大肠杆菌中表达。结果也非常好。

假设聚-Gly 缺失可能是一种改进表达的常规策略。因此鉴定了含有类似 (Gly)_n 基序(接近 N-末端，丝氨酸的下游)的其它 MenB 脂蛋白，称为 Tbp2(NMB0460)、741(NMB1870)和 983(NMB1969)：

Tbp2 ⇨ AGTbp2

```

1  MNPLVNQAA MVLVVFLLSA CLGGGGSFDL DSVDTAPRP APKYQDVFSH
51  KPQAKDQGG YGFAMRLKRR NWYPQAKEDS VKLDESUWEA TGLPDEPKEL
101 PKRQKSVIEK VETDSNNIY SSPYLKPSNH QNGNTGNGIN QPKNQAKDYE
151 NFKYVYSGWF YKHAKREFNL KVEPKSAKNG DDGYIFYHGK EPRQLPASG
201 KITYKGVWHF ATDTKGGQKF REIIQPSKSQ GDRYSGFSGD DQREYSNIOK
251 STLTDGQBGY GFTSNLEVDH HNKLTGKLI RNMANTDNNQ ATTTQYYSLE
301 AQVTGNRFNG KATATDKPQQ NSETKEHPFV SDSSSLSGGF FGPQGERLGF
351 RFLSDDQKVA VVGSAKTKDK PANGNTAAAS GGTDAASNG AAGTSSENGK
401 LITVLDVAVL KLGDKEVQKL DNFSAQAQLV VDGIPIPLP EASESGNNQA
451 NQGTNGGTAF TRKFDHTPES DKKDAQAGTQ TNGAQASNT AGDTNGKTKT

501 YEVEVCCSNL NYLKYGMLTR KNSKSAMQAG ESSSQADAKT BQVEQSMFLQ
551 GERTDEKEIP SEQWIVYRGS WYGYLANDKS TWSGNASNA TSGNRAEFTV
601 NFADKKITGT LTADNRQEAT FTIDGNLKDQ GFEGTAKTAE SGFDLQSSNT
651 TRTPKAYITD AKVQGGFYGP KAEELGGWFA YPGDKQTKNA TNASGNSSAT
701 VVFGAKRQQP VR*

```

741 ⇨ AG741

```

1  VNRTAPCCLS LTTALILTAC SSGGGGVAAD IGAGLADALT APLDHKKGL
51  QSLTLDQSVR KNEKLKLAAG GAKTYGNGD SLNTGKLEND KVSRRPDIRO
101 IEVDGQLITL ESQEFQVYKQ SHSALTAFOT EQIQDSEESG KMAKQGFRI
151 GDIAGEHTSF DKLPEGGRAT YRGTAFGSDD AGGKLTYTID PAAKQNGKI
201 EHLKSPHLNV DLAAADIKPD GKRHAVISGS VLYNQAEKGS YSLGIFGGKA
251 QEVAGSAEVK TVNGIRHIGL AAKQ*

```

983 ⇨ AG983

```

1  MRTTPTFFPK TFKPTAMALA VATTLSACLG GGGGGTSAPD FNAGGTGIGS
51  NSRATTAKSA AVSYAGIKNE MCKDRSMLCA GRDDVAVTDR DAKINAPFPN
101 LHTGDFPNPN DAYKNLINLK PAIRAGYTGR GVEVGIVDTG ESVGSIFFPE
151 LYGRKEHGYN ENYKNYTAYM RKEAPEDGGG KDIEASFDOE AVIETRAKPT
201 DIRHVKIIGH IDLVSHIIGG RSVDRPAGG LAPDATLHIM NTNDETKNEM
251 MVAAIRNANV KLGERGVRIV MNSFGTTSRA GTADLFQIAN SEEQVYRQAL
301 DYSGGDKTDE GIRLMQSDY GNLSYHKNK NMLFIFSTGN DAQAQPNFYA
351 LLPFYEKDAQ KGIITVAGVD RSGRKFYKRE YGEPGTEPLE YGSNHCGITA
401 MWCLSAPIYA SVRFTRTNPI QIAGTSFSAP IVTGTRALLL QKYPWMSNDN
451 LRTILLTTAQ DIGAVGVDSK FGWGLLDAGK AMNGPASFPF GDFATDTKOT
501 SDIAYSFRND ISGTGGLIKK GGSQQLQHN NTYTGKTIIE GGSVLVYGNV
551 KSDMRVETKG ALIYNGAASG GSLNSDGIVY LADTDQSGAN ETVHTRKSLQ
601 LDGKGTLYTR LKLLKVDGT AIIIGGLYMS ARGKAGYLN STGRVVFPLS
651 AAKIGQDYSF FTNIETDGL LASLDSVET AGSEGDTLSY YVRRGNAART
701 ASAAAHSAFA GLKHAEBQGG SNLENLMVEL DASESSATPE TVETAAADRT
751 DMPGIRPYGA TFRAAAAVQH ANAADGVRIF NSLAATVYAD STAAHADMQG
801 RRLKAVSDGL DHNGTGKLV I AQTQQDGGTW BQGGVBEKMR GSTQTVGIAA
851 KTGENTTAAA TLGMGRSTWS ENSANAKTDS ISLPAGIRHD AGDIGYLGKL
901 FSYGRYKNSI SRSTGADEHA EGSVNGTLMQ LGALGGVNVF PAATGDLTVE
951 GGLRYDLLKQ DAFAEKGSAL GWSGNSLTFE TLVLGAGLKL SQPLSDKAVL
1001 FATAGVERDL NGRDYTVTGG FTGATAATGK TGARNMPETR LVNGLGADVE
1051 FGNQWNGLAR YSYAGSKQYG NBSGRVGVGY RF*

```

Tbp2 和 741 基因本源于菌株 MC58；983 和 287 基因来源于菌株 2996。将它们克隆入 pET 载体中，并在大肠杆菌中表达而无编码它们前导肽的序列或以 ‘ Δ G 形式’(都融合于 C-端 His-标记)。在这每种情况中，发现相同的作用-在携带聚-甘氨酸段缺失的克隆中很好表达，如果在表达的蛋白质中存在甘氨酸则表达很差或不表达：

ORF	表达	纯化	细菌活性
287-His(2996)	+/-	+	+
‘287 未标记的’ 2996)	+/-	nd	nd
ΔG287-His(2996)	+	+	+
ΔG287K(2996)	+	+	+
ΔG287-His(MC58)	+	+	+
ΔG287-His(1000)	+	+	+
ΔG287-His(BZ232)	+	+	+
Tbp2-His(MC58)	+/-	nd	nd
ΔGTbp2-His(MC58)	+	+	
741-His(MC58)	+/-	nd	nd
ΔG741-His(MC58)	+	+	
983-His (2996)			
ΔG983-His (2996)	+	+	

图 13 显示了蛋白质的 SDS-PAGE。

ΔG287 和杂交体

- 制备和纯化菌株 MC58、1000 和 BZ232 的 ΔG287 蛋白质。它们都产生高 ELISA 滴定度, 且血清杀菌滴定度 > 8192。由 pET-24b 表达的 ΔG287K 在 ELISA 和血清杀菌试验中的滴定度极佳。也可在 pET-24b 中表达 ΔG287-ORF46.1K。

还将 ΔG287 直接融合于 919、953、961 和 ORF46.1 的符合读框的上游(如下显示的序列):

ΔG287-919

```

1  ATGGCTAGCC  CCGATGTTAA  ATCGGCGGAC  ACGCTGTCAA  AACC GGCCGC
51  TCCTGTTGTT  GCTGAAAAAG  AGACAGAGGT  AAAAGAAGAT  GCGCCACAGG
101  CAGGTCTCTCA  AGGACAGGGC  GCGCCATCCA  CACAAGGCAG  CCAAGATATG
151  GCGGCAGTTT  CCGCAGAAAA  TACAGGCAAT  GCGCGTGCCG  CAACAACGGA
201  CAAACCCAAA  AATGAAGACG  AGGGACCGCA  AAATGATATG  CCGCAAAATF
251  CCGCCGAATC  CGCAAAATCAA  ACAGGGAACA  ACCAAGCCGC  CGATTCTTCA
301  GATTCCGCCC  CCGCGTCAAA  CCCTGCACCT  GCGAATGGCG  GTAGCAATTT
351  TGGAAGGGTT  GATTTCGGTA  ATGGCGTTT  GATTGATGGG  CCGTCGAAA
401  ATATAACGTT  GACCCACTGT  AAAGCGGATT  CTTGTATATG  TGATAAATTA
451  TTGGATGAAG  AAGCACCGTC  AAAATCAGAA  TTTGAAAAAT  TAAATGAATC
501  TGAACGAATT  GAGAAATATA  AGAAAGATGG  GAAAAGCGAT  AAATTTACTA
551  ATTTGGTTGC  GACAGCAGTT  CAAGCTAATG  GAACFAACAA  ATATGTCATC
601  ATTTATAAAG  ACAAGTCCGC  TTCATCTTCA  TCTGCGCGAT  TCAGGCGTTC
651  TGCACGGTCG  AGGAGGTCCG  TTCTTGCCGA  GATGCCGCTA  ATCCCGTCA
701  ATCAGGCGGA  TACGCTGATT  GTCGATGGGG  AAGCGGTGAG  CCTGACGGGG
751  CATTCCGGCA  ATATCTTGGC  GCCCGAAGGG  AATTACCGGT  ATCTGACTTA
801  CGGGGCGGAA  AAATTGCCCC  GCGGATCGTA  TGCCCTCCGT  GTGCAAGGCG
851  AACC GGCAAA  AGCGGAAATG  CTTGCTGGCA  CGGCCGTGTA  CAACGGCGAA
901  GTGCTGCATT  TTCATACGGA  AAACGGCCGT  CCGTACCGGA  CTAGAGGCAG
951  GTTTGCCGCA  AAAGTCGATT  TCGGCAGCAA  ATCTGTGGAC  GGCATTATCG
1001  ACAGCGGCGA  TGATTTCAT  ATGGGTACGC  AAAAAATCAA  AGCCGCCATC

```

1051 GATGGAAACG GCTTTAAGGG GACTTGGACG GAAAATGGCG GCGGGGATGT
 1101 TTCCGGAAGG TTTTACGGCC CGGCCGGCGA GGAAGTGGCG GGAAAATACA
 1151 GCTATCGCCC GACAGATGCG GAAAAGGGCG GATTGGCGGT GTTTCGGGC
 1201 AAAAAAGAGC AGGATGGATC CGGAGGAGGA GGATGCCAAA GCAAGAGCAT
 1251 CCAAACCTTT CCGCAACCCG ACACATCCGT CATCAACGGC CCGGACCGGC
 1301 CGGTCGGCAT CCCCAGCCCC GCGGGAACGA CGGTCCGCGG CGGCGGGGCC
 1351 GTCTATACCG TTGTACCGCA CCTGTCCCTG CCCACTGGG CGGCGCAGGA
 1401 TTTCGCCAAA AGCCTGCAAT CCTTCCGCCT CGGCTGCGCC AATTTGAAAA
 1451 ACCGCCAAGG CTGGCAGGAT GTGTGCGCCC AAGCCTTTCA AACCCCGTC
 1501 CATTCTTTTC AGGCAAAACA GTTTTTTGAA CGCTATTTCA CGCCGTGGCA
 1551 GGTTCGAGGC AACGGAAGCC TTGCCGGTAC GOTTACCGGC TATTACGAGC
 1601 CGGTGCTGAA GGGCGACGAC AGGCGGACGG CACAAGCCCG CTTCCTCGATT
 1651 TACGGTATTC CCGACGATTT TATCTCCGTC CCCCTGCGTG CCGGTTTGCG
 1701 GAGCGGAAAA GCCCTGTGTC GCATCAGGCA GACGGGAAAA AACAGCGGCA
 1751 CAATCGACAA TACCGGCGGC ACACATACCG CCGACCTCTC CCGATTCCCC
 1801 ATCACC CGCG GCACAACGGC AATCAAAGGC AGGTTTGAA GAGCCGCTT
 1851 CCTCCCTTAC CACACGCGCA ACCAAATCAA CGGCGCGCGG CTTCAGCGCA
 1901 AAGCCCGCAT ACTCGGTTAC GCCGAAGACC CCGTCGAACT TTTTTTTATG
 1951 CACATCCAAG GCTCGGCGCG TCTGAAAACC CCGTCCGCA AATACATCCG
 2001 CATCGGCTAT GCGGACAAAA ACGAACATCC CTACGTTTCC ATCGGACGCT
 2051 ATATGGCGGA CAAAGGCTAC CTCAAGCTCG GGCAGACCTC GATGCAGGGC
 2101 ATCAAAGCCT ATATGCGGCA AAATCCGCAA CGCCTCGCGG AAGTTTGGG
 2151 TCAAAACCCC AGCTATATCT TTTCCGCGA GCTTGCGGGA AGCAGCAATG
 2201 ACGGTCCCGT CGGCGCACTG GGCACGCGGT TGATGGGGGA ATATGCCGGC
 2251 CGAGTCAAGC GGCATACAT TACCTTGGGC GCGCCCTTAT TTGTGCGCAC
 2301 CGCCCATCCG GTTACCGGCA AAGCCCTCAA CCGCCTGATT ATGGCGCAGG
 2351 ATACCGGCGAG CGCGATTAAA GGCGCGGTGC GCGTGGATTA TTTTGGGGGA
 2401 TACGGCGAGC AAGCCGCGCA ACTTGC CGGC AAACAGAAAA CCACGGGTTA
 2451 CGTCTGGCAG CTCTACCCA ACGGTATGAA GCCCGAATAC CGCCCGTAAC
 2501 TCGAG

1 MASPIVKSAD TLSKPAAPVV AEKETEVED APQAGSQCG APSTQGSQDM
 51 AASBAENTGN GGAATTDKPK NEDBGPQNDM PQNSAHSANQ TGMNQPADSS
 101 DAVSAPNPAP ANGGSNFGRV DLANGVLIDG PSQNTLTTCG KGDSCNGDNL
 151 LDKEAPSKSE FENLWESERI EKYKDKGSD KPTNLVATAV QANGTNKYVI
 201 IYKDKSASSS SARFRSARS RRSLEPAMPL IPVNQADTLI VDGEAVSLTG
 251 HSGNIFAPFG NYRYLTYGAE KLPGGSYALR VQGEPAKGEM LAGTAVYNGE
 301 VLHFHTENGR PYPTRGRFAA KVDGFSKSD GIIDSGDDLH MOTQKFAAI
 351 DGNQFKGTWT ENGGGDVSGR FYGPAGEEVA GKYSYRPTDA EKGFGVVFAG
 401 KKEQDGGSGG GCQSKSIQTF PQPDTSVING PDRPVGIIDP AGTTVGGGGA
 451 VYTVVPHLSL PFWAAQDFAK SLQSFRLGCA NLKNRQWQD VCAQAFQTFV
 501 HSFQAKQFFE RYFTPWQVAG NGLAGTVTG YYEPVLKDD RRTAQARFPI
 551 YGIPDDFISV FLPAGLRSQK ALVRIRQTGK NSGTIDNTGG THTADLSRFP
 601 ITARTAIKNG RFEGSRFLPY HTRNQINGGA LDGKAPILGY AEDFVELFFM
 651 HIQSGRLKLT PSQKYIRIGY ADKNEHPYVS IGRYMADEGY LKLGQTSMQG
 701 IKAYMRQNPQ RLAEVLQNP SYIFFRELAG SSNDGPVQAL GTPLMGEYAG
 751 AVDRHYITLG APLFVATAHP VTRKALNRLI MAQDTGSAIK GAURVDYFWG
 801 YGDEAGELAG KQKTGTVWQ LLPNGMKPEY RP*

AG287-953

1 ATGGCTAGCC CCGATGTAA ATCGCGGAC ACGCTGTCAA AACCGGCCGC
 51 TCCTGTTGTT GCTGAAAAAG AGACAGAGGT AAAAGAAGAT GCGCCACAGG
 101 CAGGTCTTCA AGGACAGGGC GCGCCATCCA CACAAGGCAG CCAAGATATG
 151 GCGGCACTTT CCGCAGAAAA TACAGGCAAT GCGGTTGCGG CAACAACGGA
 201 CAAACCCAAA AATGAAGACG AGGGACCGCA AAATGATATG CCGCAAAATT
 251 CCGCCGAATC CGCAAAATCAA ACAGGGAACA ACCAACC CGATTCTTCA
 301 GATTCCGCCC CCGCGTCAA CCCTGCACCT GCGAATGGCG GTAGCAATTT
 351 TGGAAAGGTT GATTTGGCTA ATGGCOTTTT GATTGATGG CCGTCGCAAA
 401 ATATAACGTT GACCCACTGT AAAGGCGATT CTGTAAATGG TGATAATTTA
 451 TTGGATGAAG AAGCACCCTC AAAATCAGAA TTTGAAAAAT TAAATGAGTC
 501 TGAACGAATT GAGAAATATA AGAAAGATGG GAAAAGCGAT AAATTTACTA
 551 ATTTGOTTGC GACAGCAGTT CAAGCTAATG GAACTAACAA ATATGTCATC
 601 ATTTATAAAG ACAAGTCCGC TTCATCTTCA TCTGCGCGAT TCAGGCGTTC
 651 TGCACGGTCG AGGAGGTCGC TTCCTGCCGA GATGCCGCTA ATCCCGTCA
 701 ATCAGGCGGA TACGCTGATT GTCGATGGG AAGCGGTCAG CCTGACGGGG
 751 CATTCCGCGA ATATCTTCGC GCCCGAAGGG AATTACCGGT ATCTGACTTA

```

801 CGGGGCGGAA AAATGCCCCG GCGGATCGTA TGCCCTCCGT GTGCAAGGCG
851 AACCGGCAAA AGGCGAAATG CTTGCTGGCA CGGCCGTGTA CAACGGCGAA
901 GTGCTGCATT TTCATACGGA AAACGGCCGT CCGTACCCGA CTAGAGGCAG
951 GTTTGCCGCA AAAGTCGATT TCGGCAGCAA ATCTGTGGAC GGCATTATCG
1001 ACAGCGGCGA TGATTTGCAT ATGGGTACGC AAAAATTCAA AGCCGCCATC
1051 GATGGAAACG GCTTTAAGGG GACTTGGACG GAAAATGGCG GCGGGGATGT
1101 TTCCGGAAGG TTTTACGGCC CGGCCGCGA GGAAGTGGCG GGAAATACA
1151 GCTATCGCCC GACAGATGCG GAAAAGGGCG GATTCGGCGT GTTTGCCGGC
1201 AAAAAAGAGC AGGATGGATC CGGAGGAGGA GGAGCCACCT ACAAAOTGGA
1251 CGAATATCAC GCCAACGCCC GTTTCGCCAT CGACCATTC AACACCAGCA
1301 CCAACGTCGG CGGTTTTCAC GGTCTGACCG GTTCCGTGCA GTTCGACCAA
1351 GCAAAACGCG ACGGTAAAAAT CGACATCACC ATCCCCGTG CCAACCTGCA
1401 AAGCGGTTTC CAACACTTTA CCGACCACCT GAAATCAGCC GACATCTTCG
1451 ATGCCGCCCA ATATCCGGAC ATCCGCTTTG TTTCCACCAA ATTCAACTTC
1501 AACGGCAAAA AACTGGTTTC CGTTGACGGC AACCTGACCA TGCACGGCAA
1551 AACCGCCCCC GTCAAACTCA AAGCCGAAAA ATTCAACTGC TACCAGAGCC
1601 CGATGGCGAA AACCAGAGTT TCGGCGCGCG ACTTCAGCAC CACCATCGAC
1651 CGCACCAAAAT GGGGCGTGGG CTACCTCGTT AACGTTGGTA TGACCAAAAG
1701 CGTCCGCATC GACATCCAAA TCGAGGCAGC CAAACAATAA CTCGAG

1 MASPDVKSAD TLSKPAAPVV AEKETEVKED APQAGSQGG APSTQGSQDM
51 AAVSAENTGN GGAATTDKPK NEDEGPQNDM PQNSAESANQ TGNHNPADSS
101 DSAPASNPAP ANGGSNFRV DLANGVLIDG PSQMITLTHC KGDSCNGDNL
151 LDEKAPSKSE FENLHESERI EKYKDKGSD KFTMLVATAV QANGTNYKVI
201 IYKKSASSS SARFRSARS RRLPAEMPL IPVNQADTLI VDGRAVSLTG
251 HSGNIFAPBG NYRYLTYGAE KLPQGSYALR VQGEPAKQEM LAGTAVYNGE
301 VLHFTENGFR PYPTGRFAA KVDGSKSVD GIIDSGDDLH MOTQKFKAAI
351 DNGFVKGTWT ENGGGDVSGR FYGPAGREVA GKYSYRPTDA EKGFGVVFAG
401 KKEQDGSQQG GATYKVDYH ANARFAIDHF NTSTNVGGFY GLTGSVEFDQ
451 AKRDGKIDIT IPVANLQSGS QHFTDHLKSA DIFDAAQYPD IRFVSTKFNF
501 NGKILVSDG NLTMHGKTAP VKLKAKFPNC YQSPMAKTEV CGGDFSTTID
551 RTKWGVVDYLV NVGMTKSVRI DIQIEAAQ*

```

AG287-961

```

1 ATGGCTAGCC CCGATGTTAA ATCGGCGGAC ACGCTGTCAA AACCGGCCGC
51 TCCTGTTGTT GCTGAAAAAG AGACAGAGGT AAAAGAAGAT GCGCCACAGG
101 CAGGTTCTCA AGGACAGGGC GCGCCATCCA CACAAGGCAG CCAAGATATG
151 GCGCAGTTT CCGCAGAAAA TACAGGCAAT GCGCGTGGCG CAACAACGGA
201 CAAACCCAAA AATGAAGACG AGGACCCGCA AAATGATATG CCGCAAAATT
251 CCGCCGAATC CGCAAAATCA ACAGGGAACA ACCAACCAGC CGATTCTTCA
301 GATTCCGCCC CCGCGTCAA CCTGACACT GCGAATGGCG GTAGCAATTT
351 TGGAAGGGTT GATTTGGCTA ATGGCGTTT GATTGATGGG CCGTCGCAAA
401 ATATAACGTT GACCCACTGT AAGGCGGATT CTTGTAATGG TGATAATTTA
451 TTGGATGAAG AAGCACCCTC AAAATCAGAA TTTGAAAATT TAAATGAGTC
501 TGAACGAATT GAGAAATATA AGAAAGATGG GAAAAGCGAT AAATTTACTA
551 ATTTGGTTGC GACAGCAGTT CAAGCTAATG GAACTAACAA ATATGTCATC
601 ATTTATAAAG ACAAGTCCGC TTCATCTTCA TCTGCCGAT TCAGGCGTTC
651 TGCACGGTCG AGGAGGTGCG TTCTGCCGA GATGCCGCTA ATCCCCGTCA
701 ATCAGGCGGA TACGCTGATT GTCGATGGGG AAGCGGTCAG CCTGACGGGG
751 CATTCGGGCA ATATCTTGGC GCCGAAGGG AATTACCGGT ATCTGACTTA
801 CCGGGCGGAA AAATTGCCCG GCGGATCGTA TGCCCTCCGT GTGCAAGGCG
851 AACCGGCAAA AGGCGAAATG CTTGCTGGCA CGGCCGTGTA CAACGGCGAA
901 GTGCTGCATT TTCATACGGA AAACGGCCGT CCGTACCCGA CTAGAGGCAG
951 GTTTGCCGCA AAAGTCGATT TCGGCAGCAA ATCTGTGGAC GGCATTATCG
1001 ACAGCGGCGA TGATTTGCAT ATGGGTACGC AAAAATTCAA AGCCGCCATC
1051 GATGGAAACG GCTTTAAGGG GACTTGGACG GAAAATGGCG GCGGGGATGT
1101 TTCCGGAAGG TTTTACGGCC CGGCCGCGA GGAAGTGGCG GGAAATACA
1151 GCTATCGCCC GACAGATGCG GAAAAGGGCG GATTCGGCGT GTTTGCCGGC
1201 AAAAAAGAGC AGGATGGATC CGGAGGAGGA GGAGCCACAA ACGACGACGA
1251 TGTAAAAAAG GCTGCCACTG TGCCCATTCG TGCTGCCCTAC AACAAATGGCC
1301 AAGAAATCAA CGGTTTCAA GCTGGAGAGA CCATCTACGA CATGATGAA
1351 GACGGCACAA TTACCAAAAA AGACGCAACT GCAGCCGATG TTGAAGCCGA
1401 CGACTTTAAA GGTCTGGGTC TGAAAAAAGT CGTGACTAAC CTGACCAAAA
1451 CCGTCAATGA AAACAAACAA AACGTCGATG CCAAAGTAAA AGCTGCAGAA
1501 TCTGAAATAG AAAAGTTAAC AACCAAGTTA GCAGACACTG ATGCCGCTTT
1551 AGCAGATACT GATCCCGCTC TGGATGCAAC CACCAACGCC TTGAATAAAT

```

```

1601 TGGGAGAAAA TATAACGACA TTTGCTGAAG AGACTAAGAC AAATATCGTA
1651 AAAATTTGATG AAAAAATTAGA AGCCGTGGCT GATACCGTGG ACAAGCATGC
1701 CGAAGCATTC AACGATATCG CCGATTCATT GGATGAAACC AACACTAAGG
1751 CAGACGAAGC CGTCAAAACC GCCAATGAAG CCAAACAGAC GGCCGAAGAA
1801 ACCAAACAAA ACGTCGATGC CAAAGTAAAA GCTGCAGAAA CTGCAGCAGG
1851 CAAAGCCGAA GCTGCCGCTG GCACAGCTAA TACTGCAGCC GACAAGGCCG
1901 AAGCTGTGCG TGCAAAAGTT ACCGACATCA AAGCTGATAT CGCTACGAAC
1951 AAAGATAATA TTGCTAAAAA AGCAAACAGT GCCGACGTGT ACACCAGAGA
2001 AGAGTCTGAC AGCAAATTTG TCAGAATTGA TGGTCTGAAC GCTACTACCG
2051 AAAAATTGGA CACACGCTTG GCTTCTGCTG AAAAAATCCAT TGCCGATCAC
2101 GATACTCGCC TGAACGGTTT GGATAAAACA GTGTCAAGAC TGCGCAAAGA
2151 AACC CGCCAA GGCCTTGCAG AACAGCCGC GCTCTCCGGT CTGTTCCAAC
2201 CTTACAACGT GGGTCGGTTC AATGTAACGG CTGCAGTCGG CGGCTACAAA
2251 TCCGAATCGG CAGTCGCCAT CGGTACCGGC TTCCGCTTTA CCGAAAACCT
2301 TGCCGCCAAA GCAGGCGTGG CAGTCGCCAC TTCGTCCGGT TCTTCCGCAG
2351 CCTACCATGT CGGCGTCAAT TACGAGTGGT AACTCGAG

```

```

1 MASPDVKSAD TLSKPAAPVV AEKETEVKED APQAGSQGG APSTQGSQDM
51 AAVSAEMTGN GGAATTDKPK NEDEBPQNDM PQNSAESANQ TGNWQPADSS
101 DSAPASNAP ANGGSNFGRV DLANGVLIDG PSQNTITLTHC KGDSCNGDNL
151 LDEEAPSKSE FENLNESERI EKYKKDGRSD KFTNLVATAV QANGTKYVI
201 IYKDKSASSS SARFRESARS RRLPAEMPL IFVWQADTLI VDGEAVSLTG
251 HSGNIFAPEG NYRYLYGAE KLPGGSYALR VQGEPAKGEM LAGTAVYNGE
301 VLHFTENGRI PYPTGRFAA KVDFGSKSVD GIIDSGDDLH MGTQKPKAAI
351 DGNFGKGTWT ENGGGDVSGR FYGPRAGEVA GKYSYRPTDA EKGGFQVFPAG
401 KKEQDQSGGG GATNDDDVKK AATVAIAAAY MNGQELNGFK AGETIYDIDE
451 DGTITTKDAT AADVEADDFK GLGLKKVVTN LKTVNENKQ NYDAKVKAAE
501 SEIEKLTKL ADTDAALADT DAALDATNA LNKLGENTTT FARETKTNIV
551 KIDEKLEAVA DTVDEHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEK
601 TKQNVDAKVK AASTAAGKAE AAAGTANTAA DEAEVAARKV TDIEKADIATN
651 KDNIAKANS ADVYTRERSD SKFVRIDGLN ATTEKLDTRL ASAEKSTADH
701 DTRLNGLDKT VSDLRETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK
751 SESAVAIGTG FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEW*

```

	ELISA	杀菌
Δ G287-953-His	3834	65536
Δ G287-961-His	108627	65536

将抗体针对杂交蛋白产生的杀菌效力(同源菌株)与针对 919 和 ORF46.1 的组分抗原(使用 287-GST)的简单化合物产生的抗体相比:

	与 287 的混合物	与 Δ G287 的杂交体
919	32000	128000
ORF46.1	128	16000

还获得了针对异源 MenB 菌株和血清型 A 和 C 的杀菌活性的数据:

菌株	919		ORF46.1	
	混合物	杂交体	混合物	杂交体
NGH38	1024	32000	-	16384
MC58	512	8192	-	512
BZ232	512	512	-	-
MenA (F6124)	512	32000	-	8192
MenC (C11)	>2048	>2048	-	-
MenC (BZ133)	>4096	64000	-	8192

因此, 在与 Δ G287(在 N-末端)的杂交蛋白质比简单的与 Δ G287-ORF46.1的混合物在免疫学上更优良, 针对异源菌株尤其有效。也可在 pET-24b 中表达 Δ G287-ORF46.1。

用新西兰菌株 394/98 而非 2996 制备相同的杂交蛋白:

AG287ME-919

```

1   ATGGCTAGCC CCGATGTCAA GTCGGCGGAC ACGCTGTCAA AACCTGCCGC
51  CCCTGTTGTT TCTGAAAAAG AGACAGAGGC AAAGGAAGAT GCCCCACAGG
101 CAGGTTCTCA AGGACAGGGC GCGCCATCCG CACAGGCGG TCAAGATATG
151 GCGGCGGTTT CGGAAGAAAA TACAGGCAAT GCGCGTCCGG CAGCAACGGA
201 CAAACCCAAA AATGAAGACG AGGGGGCGCA AATGATATG CCGCAAAATG
251 CCGCCGATAC AGATAGTTTG ACACCGAATC ACACCCCGGC TTCGAATATG
301 CCGGCCGGA AATATGAAAA CCAAGCACCG GATGCCGGGG AATCGAGCA
351 GCCGGCAAAC CAACCGGATA TGGCAAATAC GCGCGACGGA ATGCAGGGTG
401 ACGATCCGTC GGCAGGCGGG GAAAAATGCC GCAATACGGC TGCCCAAGGT
451 ACAAAATCAAG CCGAAAACAA TCAAACCGCC GGTTCCTCAA ATCCTGCCTC
501 TTCAACCAAT CCTAGCGCCA CGAATAGCGG TGGTGATTTT GGAAGGACGA
551 ACGTGGGCAA TTCTGTTGTG ATTGACGGGC CGTCGCAAAA TATAACGTTG
601 ACCCACTGTA AAGGCGATTC TTGTAGTGGC AATAATTTCT TGGATGAAGA
651 AGTACAGCTA AAATCAGAAT TTGAAAAATT AAGTGATGCA GACAAAAATA
701 GTAATTACAA GAAAGATGGG AAGAATGACG GGAAGAAATG TAAATTGTTC
751 GGTTTGGTTG CCGATAGTGT GCAGATGAAG GGAATCAATC AATATATTAT
801 CTTTTATAAA CCTAAACCCA CTTCATTGTC GCGATTAGG CGTCTGTCAC
851 GGTGAGGGCG GTCGCTTCCG GCCGAGATGC CGCTGATTCC CGTCAATCAG
901 GCGGATACGC TGATTGTGGA TGGGGAAGCG GTCAGCCTGA CCGGGCATTC
951 CGGCAATATC TTCGCGCCCG AAGGGAAATTA CCGGTATCTG ACTTACGGGG
1001 CGGAAAAATT GCCCGGCGGA TCGTATGCCC TCCGTGTTCA AGGCGAACCT
1051 TCAAAAGGCG AAATGCTCGC GGGCACGGCA GTGTACAACG GCQAAGTGCT
1101 GCATTTTCAT ACGGAAAACG GCCGTCCGTC CCCGTCCAGA GCGAGGTTTG
1151 CCGCAAAAGT CGATTTCGCG AGCAAACTG TGGACGGCAT TATCGACAGC
1201 GCGGATGOTT TGCATATGGG TACGCAAAAA TTCAAAGCCG CCATCGATGG
1251 AAACGGCTTT AAGGGGACTT GACCGGAAAA TGGCGGCGGG GATGTTTCCG
1301 GAAAGTTTFA CGCCCGGCC GCGGAGGAAG TGGCGGAAA ATACAGTAT
1351 CGCCCAACAG ATGCGGAAAA GGGCGGATTC GCGGTGTTG CCGGCAAAAA
1401 AGAGCAGGAT GGATCCGAG GAGGAGGATG CCMAAGCAAG AGCATCCAAA
1451 CCTTTCCGCA ACCCGACACA TCCGTCAATCA ACGGCCCGGA CCGGCCGGTC
1501 GGCATCCCGG ACCCGCGCG AACGACGGTC GCGGCGGGG GGGCCGTCTA
1551 TACCGTTGTA CCGCACCTGT CCTGCCCA CTGGGCGGG CAGGATTTTG
1601 CCAAAAGCCT GCAATCCTTC CGCCTCGGCT GCGCCAATTT GAAAAACCGC
1651 CAAGGCTGGC AGGATGTGTG CGCCCAAGCC TTTCAAAACC CGTCCATTC
1701 CTTTCAGGCA AAACAGTTT TTGAACGCTA TTTCACGCG TGGCAGGTTG
1751 CAGGCAACGG AAGCCTTGCC GTTACGGTTA CCGGCTATTA CGAGCCGGTG
1801 CTGAAGGGCG ACGACAGGCG GACGGCACAA GCCCGCTTCC CGATTACGG
1851 TATTCGCGAC GATTTTATCT CGTCCCGCT GCTGCGGGT TTGCGGAGCG
1901 GAAAAGCCCT TGTCCGCATC AGGCAGACGG GAAAAACAG CCGCACATC
1951 GACAATACCG GCGGCACACA TACCGCGAC CTCTCCGAT TCCCCATCAC
2001 CGCGCGCACA ACGGCAATCA AAGGCAGGTT TGAAGGAAGC CGCTTCCTCC

```

```

2051 CCTACCACAC GCGCAACCAA ATCAACGGCG GCGCGCTTGA CGGCAAAGCC
2101 CCGATACTCG GTTACGCCGA AGACCCCGTC GAACCTTTTTT TTATGCACAT
2151 CCAAGGCTCG GGCCGCTCTGA AAACCCCGTC CGGCAAATAC ATCCGCATCG
2201 GCTATGCCGA CAAAAACGAA CATCCCTACG TTTCCATCGG ACGCTATATG
2251 GCGGACAAAG GCTACCTCAA GCTCGGGCAG ACCTCGATGC AGGCATCAA
2301 AGCCTATATG CGGCAAATC CGCAACGCCT CGCCGAAGTT TTGGGTCAAA
2351 ACCCCAGCTA TATCTTTTTC CGCGAGCTTG CCGGAAGCAG CAATGACGGT
2401 CCCGTCGGCG CACTGGGCAC GCCGTGTGATG GGGGAATATG CCGCGCAGT
2451 CGACCCGCAC TACATTACCT TGGGCGCGCC CTATTTTGTG GCCACCGCCC
2501 ATCCGGTTAC CCGCAAAGCC CTCACCGGCC TGATTATGGC GCAGGATACC
2551 GGCAGCGCGA TTAAAGCGCG GGTGCGCGTG GATTATTTT GGGGATACGG
2601 CGACGAAGCC GCGCAACTTG CCGGCAACA GAAAACCACG GGTACGTCT
2651 GGCAGCTCCT ACCCAACGGT ATGAAGCCCG AATACCGCCC GTAAAGCTT

```

```

1 MASPDVKSAD TLSKPAAPVV SEKETEAKED APQAGSQGG APSAQGGQDM
51 AAVSEENTGN GGAAATDKPK NEDEGACNDM PQMAADTDSL TPNHTPASNM
101 PAGNMENQAP DAGESBQPAN QPDMANTADG MQGDDPSAGG ENAGNTAAQG
151 TNQAKENQTA GSQNPASSTN PSATNSGGDF GRINVGNFV IDGPSQNTIL
201 THCKGDCSCG NNFLDEEVQL KSEFEKLSDA DKISNYKKDG KNDGKNDKRV
251 GLVADSVQMK GINQYIIFYK PKPTSFARFR RSARSRRSLP AEMPLIFVNG
301 ADTLIVDGEA VSLTGHSGNI FAPEGNYRYL TYGAEKLPFG SYALRVQGEF
351 SKGEMLAGTA VYNGEVLHFH TENGRPSPSR GRFAAKVDFG SKSVDDIIDS
401 GDGLHMTQK FKAIDGNF KGTWTEGNGG DVSQKFYGA GEEVAGKYSY
451 RPTDAEKGGF GVFAGEKEQD GSGGGGCGSK SIQTFFQPD SVINGPDRPV
501 GIPDPAGTTV GGGGAVYTVV PHLSLPHMA QDFAKSLQSF RLGCAMLKNR
551 QGWQDVCAQA FQTFVHSPQA KQFFERYFTF WQVAGNGSLA GTVTGYEFPV
601 LKGDDRRTAQ ARFFIYGIPD DFISVFLPAG LRSGKALVRI RQTGKNSGTI
651 DNTGGTHTAD LSRFPITART TAIKGRFEGS RFLPYHTRNQ INGGALDQKA
701 PILGYAEDFV ELFFMHIQGS GRUKTPSGKY IRIGYADKNE HPYVSIGRYM
751 ADKGYLKLQ TSMQGIKAYM RQNPORLAEV LGQMFPSYIF RELAGSSNDG
801 FVGALQTFPL GEYAGAVDRH YTTLGAPLFV ATAHPVTRKA LNRLTMAQDT
851 GSAIKGAVRV DYFWGYGDEA GELAGKQKTT GYVWQLLPNG MKPEYRP*

```

AG287NZ-953

```

1 ATGGCTAGCC CCGATGTCAA GTCGGCGGAC ACGCTGTCAA AACCTGCCGC
51 CCCTGTTGTT TCTGAAAAAG AGACAGAGGC AAAGGAAGAT GCGCCACAGG
101 CAGGTTCTCA AGGACAGGGC GCGCCATCCG CACAAGGCGG TCAAGATATG
151 GCGGCGGTTT CGGAAGAAAA TACAGGCAAT GCGCGTGCAG CAGCAACGGA
201 CAAACCCAAA AATGAAGACG AGGGGGCGCA AATGATATG CCGCAAAATG
251 CCGCCGATAC AGATAGTTTG ACACCGAATC ACACCCCGGC TTCGAATATG
301 CCGGCCGGAA ATATGAAAA CCAAGCACCG GATGCCGGGG AATCGGAGCA
351 GCCGGCAAAC CAACCGGATA TGGCAAATAC GCGCGACGGA ATGCAGGOTG
401 ACGATCCGTC GGCAGGCGGG GAAATGCCG GCAATACGGC TGCCCAAGGT
451 ACAATCAAG CCGAAAACAA TCAAACCGCC GGTCTCTCAA ATCTGCGCTC
501 TTCAACCAAT CCTAGCGCCA CGAATAGCGG TGGTGATTTT GGAAGGACGA
551 ACGTGGGCAA TTCTGTTGTG ATTGACGGGC CGTCGCAAAA TATAACGTTG
601 ACCCACTGTA AAGCGGATTC TTGTAGTGGC AATAATTTCT TGGATGAAGA
651 AGTACAGCTA AATCAGAAAT TTGAAAAATT AAGTGATGCA GACAAAATAA
701 GTAATTACAA GAAAGATGGG AAGAATGACG GGAAGAATGA TAAATTTGTC
751 GGTTTGGTTG CCGATAGTGT GCAGATGAAG GGAATCAATC AATATATTAT
801 CTTTATAAAA CCTAAACCCA CTTCAATTGC GCGATTTAGG CGTCTGCAAC
851 GGTGAGGGCG GTCGCTTCCG GCCGAGATGC CGCTGATTCC CGTCAATCAG
901 GCGGATACGC TGATTGTCTGA TGGGGAAGCG GTCAGCCTGA CCGGGCATTC
951 CCGCAATATC TTCGCGCCCG AAGGGAATTA CCGGTATCTG ACTTACGGGG
1001 CGGAAAAATT GCCCGCGGGA TCGTATGCC TCCGTGTTCA AGCGAAACCT
1051 TCAAAAAGCG AAATGCTCGC GGCACGGCA GTGTACAAAG CGGAAGTGCT
1101 GCATTTTCAT ACGGAAAACG GCCGTCCGTC CCCGTCCAGA GGCAGGTTTG
1151 CGCAAAAGT CGATTTCGGC AGCAAATCTG TGGACGGCAT TATCGACAGC
1201 GCGGATGGTT TGCATATGGG TACGCAAAAA TTCAAAGCCG CCATCGATGG
1251 AAACGGCTTT AAGGGGACTT GGACGGAAAA TGGCGGGGGG GATGTTTCCG
1301 GAAAGTTTTA CGGCCCGGCC GCGAGGGAAG TGGCGGAAAA ATACAGCTAT
1351 CGCCCAACAG ATGCGGAAAA GGGCGGATTC GGCCTGTTTG CCGCAAAAA
1401 AGAGCAGGAT GGATCCGGAG GAGGAGGAGC CACCTACAAA GTGGACGAAT
1451 ATCAGCCCAA CGCCCGTTTC GCCATCGACC ATTTCAACAC CAGCACCAAC
1501 GTCGGCGGTT TTTACGGTCT GACCGGTTCC GTCGAGTTCC ACCAAGCAAA
1551 ACGGACGGT AAAATCGACA TCACCATCCC CGTTGCCAAC CTGCAAGCCG

```

```

1601 GTTCGCAACA CTTTACCGAC CACCTGAAAT CAGCCGACAT CTTCGATGCC
1651 GCCCAATATC CGGACATCCG CTTTGTTCCT ACCAAATTC AACTCAACGG
1701 CAAAAAAGCT GTTTCGGTTG ACGGCAACCT GACCATGCAC GGCAAAACCG
1751 CCCCCGTCAA ACTCAAAGCC GAAAAATTC AACTGCTACCA AAGCCCGATG
1801 GCGAAAACCG AAGTTTGCGG CGGCGACTTC AGCACCACCA TCGACCGCAC
1851 CAAATGGGGC GTGGACTACC TCGTTAACGT TGGTATGACC AAAAGCGTCC
1901 GCATCGACAT CCAAATCGAG GCAGCCAAAC AATAAAAGCT T

```

```

1 MASPDVKSAD TSKPAAPVV SEKETEKED APQAGSQGG APSAQGGQDM
51 AAVSEENIGN GGAAATDKPK NEDEGAQNDM PQNAADTDSL TPNHTPASNM
101 PAGNMENQAP DAGESEQPAN QPDMANTADG MQGDDPSAGG ENAGETAAQG
151 TNQAENNQTA GSQNPASSTN PSATNSGGDF GRTNVGNSVV IDGPSQMITL
201 THCKGDCSCG NMFLDEEVQL KSEFEKLSDA DKISNYKKDG KNDGKNDKDFV
251 GLVADSVQMK GINQYIIFYK PKPTSFARFR RSARSRRSLP AEMPLIPVNO
301 ADTLIVDGEA VSLTGHSGNI FAPEGNYRYL TYGAELKPGG SYALEVQGEF
351 SKGEMLAGTA VYNGEVLHFF TENGRPSPSR GRFAAKVDFG SKSVDDGIIDS
401 GDGLHMTQK PKAALIDNGF KGTWTENGCG DVSCKFYGPA GEEVAGKYSY
451 RPTDAEKGGF GVFAGKKEQD GSGGGGATYK VDEYHANARF AIDHFTSTN
501 VGGFYGLTGS VEFDAQKRDG KIDITIPVAN LQSGSGHFTD HLKSDIFDA
551 AQYPDIRFVS TKFNFNKEL VSDGNLTMH GKTAPVKLKA EKFNCYQSPM
601 AKTEVCGGDF STTIDRTKMG VDYLNVNGMT KSVRIDQIE AAKQ*

```

AG287NE-961

```

1 ATGGCTAGCC CCGATGTCAA GTCGCGGAC ACGCTGTCAA AACCTGCCGC
51 CCCTGTTGTT TCTGAAAAG AGACAGAGGC AAGGAGAGAT GCGCCACAGG
101 CAGGTTCTCA AGGACAGGC GCGCATCCG CACAAGCGG TCAAGATATG
151 GCGCGGTTT CCGAAGAAA TACAGGCAAT GCGCGTCCG CAGCAACGGA
201 CAAACCCAAA AATGAAGACG AGGGGGCGCA AAATGATATG CCGCAAAATG
251 CCGCCGATAC AGATAGTTTG ACACCGAATC ACACCCCGG TCGAATATG
301 CCGCCGGAAT ATATGGAAA CCAAGCACCG GATGCCGGG AATCGGAGCA
351 GCCGCAAAC CAACCGGATA TGGCAAATAC GCGGACGGA ATGCAGGGTG
401 ACGATCCGTC GGCAGGCGG GAAATGCCG GCAATACGGC TGCCCAAGGT
451 ACAATCAAG CCGAAAACAA TCAAACCGCC GGTTCCTCAA ATCCTGCCTC
501 TTCAACCAAT CCTAGCGCCA CGAATAGCG TGGTGATTTT GGAAGGACGA
551 ACGTGGGCAA TTCTGTTGTG ATTGACGGG CGTGCGAAAA TATAACGTTG
601 ACCCACTGTA AAGGCGATTC TTGTAGTGG AATAATTTCT TGGATGAAGA
651 AGTACAGCTA AATCAGAAT TTGAAAAAT AAGTATGCA GACAAAATAA
701 GTAATTACAA AAGAATGGG AAGAATGAC GGAAGATGA TAAATTTGTC
751 GGTFTGGTTG CCGATAGTGT GCAGATGAAG GGAATCAATC AATATATTAT
801 CTTTATATAA CCTAAACCCA CTTCAATTGC GCGATTTAGG CGTTCTGCAC
851 GGTGAGGGCG GTCGCTTCCG GCCGAGATGC CGCTGATTC CGTCAATCAG
901 GCGGATACGC TGATTGTGCA TGGGGAAGCG GTCAGCCTGA CCGGGCATTC
951 CCGCAATATC TTCGCGCCCG AAGGGAATTA CCGGTATCTG ACTTACGGGG
1001 CCGAATAATT GCCCGCGGA TCGTATGCC TCCGTGTTCA AGGCGAACCT
1051 TCAAAGGCG AAATGCTGCG GGCACGGCA GTGTACAACG GCGAAGTGCT
1101 GCAATTTTCAT ACGGAAAACG GCCGTCCGTC CCCGTCCAGA GGCAGGTTTG
1151 CCGCAAAAGT CGATTTCCGC AGCAAATCTG TGGACGGCAT TATCGACAGC
1201 GCGGATGGTT TGCATATGGG TACGCAAAA TTCAAAGCCG CCATCGATGG
1251 AAACGGCTTT AAGGGGACTT GGACGGAAAA TGGCGGCGGG GATGTTTCCG
1301 GAAAGTTTTA CCGCCCGGCC GCGAGGAAG TGGCGGGAAT ATACAGCTAT
1351 CGCCCAACAG ATGCGGAAA GCGCGGATTC GCGGTGTTTG CCGCAAAAA
1401 AGAGCAGGAT GGATCCGGAG GAGGAGGAGC CACAAACGAC GACGATGTTA
1451 AAAAGGCTGC CACTGTGGCC ATGCTGCTG CCTACAACA TGGCCAAGAA
1501 ATCAACGGTT TCAAAGCTGG AGAGACCATC TACGACATTG ATGAAGACGG
1551 CACAATTACC AAAAAAGACG CAACTGCAGC CGATGTTGAA GCGACGACT
1601 TTAAGGTCT GGTCTGAAA AAGTCTGTA CTAACCTGAC CAAAACGTC
1651 AATGAAAACA AACAAAACGT CGATGCCAAA GTAAAGCTG CAGAATCTGA
1701 AATAGAAAAG TTAACAACCA AGTTAGCAGA CACTGATGCC GCTTTAGCAG
1751 ATACTGATGC CGCTCTGGAT GCAACCACCA ACGCTTGAA TAAATTGGGA
1801 GAAATATATA CGACATTTGC TGAAGAGACT AAGACAAATA TCGTAAAAAT
1851 TGATGAAAAA TTAGAAGCCG TGGCTGATAC CGTCGACAAG CATGCCGAAG
1901 CATTCACGTA TATCGCCGAT TCATTGGATG AAACCAACAC TAAGGCAGAC
1951 GAAGCCGTCA AAACCGCCAA TGAAGCCAAA CAGACGGCCG AAGAAACCAA
2001 ACAAAACGTC GATGCCAAG TAAAAGCTGC AGAACTGCA GCAGGCAAG
2051 CCGAAGCTGC CGCTGGCACA GCTAATACTG CAGCCGACAA GGCCGAAGCT
2101 GTCGCTGCAA AAGTTACCGA CATCAAAGCT GATATCGCTA CGAACAAAGA

```

```

2151 TAATATTGCT AAAAAAGCAA ACAGTGCCGA CGTGACACC AGAGAAGAGT
2201 CTGACAGCAA ATTTGTCAGA ATTGATGGTC TGAACGCTAC TACCGAAAAA
2251 TTGGACACAC GCTTGGCTTC TGCTGAAAAA TCCATTGCCG ATCAGGATAC
2301 TCGCCTGAAC GGTTTGGATA AAACAGTGTC AGACCTGCCG AAAGAAACCC
2351 GCCAAGGCCT TGCAGAACAA GCCGCGCTCT CCGGTCTGTT CCAACCTTAC
2401 AACGTGGGTC GGTTCATGTT AACGGCTGCA GTCGGCGGCT ACAATCCGA
2451 ATCGGCAGTC GCCATCGGTA CCGGCTTCCG CTTTACCGAA AACTTTGCCG
2501 CCAAAGCAGG COTGGCAGTC GGCACCTCGT CCGGTTCTTC CGCAGCCTAC
2551 CATGTCGGCG TCAATTACGA GTGGTAAAG CTT

1 MASFDVKSAD TLSKPAAFV SEKETEKED APQAGSQGG APSAQGGQDM
51 AAVSEBENTGN GGAATDKPK NEDEGAQNDM PQNAADTDSL TPNHTPASNM
101 PAGNMENQAP DAGESEBPAN QPDMANTADG MQGDDPSAGG ENAGNTAAQG
151 TNQAENQTA GSQNPASSTN PSATNSGGDF GRTNVGNHVV IDGPSQMITL
201 THCKGDCSG NNFLDESVQL KSEFEKLSDA DKLSHYKKDG KNDGKNDKPV
251 GLVADSVQMK GINQYILIFYK PKPTSFARFR RSARSRRSLP ABMPLIFVNO
301 ADTLIVDGEA VSLTGHSGNI FAPEGNYRYL TYGAELPGG SYALERVQGEF
351 SKGEMLAGTA VYNGEVLHFF TENGEFSPSR GRFAAKVDFG SKSVDGIIDS
401 GDGLHMGTQK FKAALDQNGF KGTWTENGCG DVBGKPYGPA GKEVAGKYSY
451 RPTDAKGGF GVFAGKKEQD GSGGGGATND DDVKAATVA IAAAYNNGQE
501 INGFKAGETI YDIDEDGTTT EKDATAADVE ADDPKGLGLK KVVYTNLTKTV
551 NENKQNVDAK VKAABSEIEK LTTKLADTDA ALADTDALD ATTMALEKLG
601 ENITTFABST KTNIVKIDRK LEAVADTVDK HAEAFNDIAD SLDSNTKAD
651 EAVTANEAK QTAETKQNV DAKVKAETA AGKABAAAGT ANTAADKAEA
701 VAAKVTDIKA DIATNKDNTA KKANSADVYT REESDSKFVR IDGLMATTEK
751 LDTRLASA EK SIADHDTRLN GLDKTVSDLR KETRQGLAEQ AALSGLFQFY
801 NVGRFNVTA VGGYKESAV AIGTGFRFTE NFAAKAGVAV GTSSGSSAAY
851 HVGVNYEW*

```

△G983 和杂交体

测定了针对各种菌株(包括同源 2996 菌株)应答△G983(His-融合体)产生的杀菌滴定度:

	2996	NGH38	BZ133
△G983	512	128	128

5 △G983 还可以与 ORF46.1、741、961 或 961c(在其 C-末端)杂交的杂交体表达:

AG983-ORF46.1

```

1 ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
51 CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
101 AGAACGAAT GTGCAAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC
151 GTTGGCGGTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT
201 GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATAACAAG AATTTGATCA
251 ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
301 GGTATCGTCG ACACAGGCGA ATCCGTCCGC AGCATATCCT TTCCCGAAGT
351 GTATGGCAGA AAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
401 CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
451 GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
501 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTGTGTC TCCCATATTA
551 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TCGCCCCGAT
601 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
651 GGTTCAGGCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
701 GCATCGTCAA TAACAGTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
751 CTTTTCAAA TAGCCAAATC GGAGGAGCAG TACCGCCAA GCTTGCTCGA
801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA
851 CGGATTACGG CAACCTGTCC TACCACATCC GTAATAAAA CATGCTTTTC
901 ATCTTTTCGA CAGGCAATGA CGCACAAGCT CAGCCCAACA CATATGCCCT
951 ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001 GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTCGCGAA TTACTGCCAT
1101 GTGGTGCCCTG TCGCACCCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA

```

```

1151 ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC
1201 ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
1251 GCGTACCACG TTGCTGACGA CCGCTCAGGA CATCGGTGCA GTCGGCGTGG
1301 ACAGCAAGTT CCGCTGGGGA CTGCTGGATG CCGGTAAGGC CATGAACGGA
1351 CCCGCTCCTT TTCCGTTCCG CCACTTTACC GCGGATACGA AAGGTACATC
1401 CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GCGCGCTGA
1451 TCAAAAAAGG CCGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501 GGCAAAACCA TTATCGAAGG CGGTTTCGTC GTGTTGTACG GCAACAACAA
1551 ATCGGATATG CCGCTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGCGCG
1601 CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCATCTT GGCAGATACC
1651 GACCAATCCG GCGCAAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT
1701 GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG
1751 ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG
1801 GGGCGAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCTT TCCTGAGTGC
1851 CGCCAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
1901 GCGGCTGCTT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGAA
1951 GCGGACACGC TGTCCCTATTA TGTCCGTCGC GGCATATCGG CACGGACTGC
2001 TTCGGCAGCG GCACATTCGG CCGCCGCGCG TCTGAAACAC GCCGTAGAAC
2051 AGGCGGGCAG CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101 TCATCCGCAA CACCCGAGAC GGTGAAACT GCGCGAGCCG ACCGACAGA
2151 TATGCCGGGC ATCCGCCCCC ACGGCGCAAC TTTCCGCGCA GCGGCGCCG
2201 TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
2251 GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TCGAGGAGC
2301 CCGCTCGAAA GCGGTATCGG ACGGTTTGGG CCAACAACGG ACGGGTCTGC
2351 GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
2401 GTTGAAGGCA AAATGCCCGG CAGTACCCAA ACCGTCCGCA TTGCCCGGAA
2451 AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGACGCA
2501 CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTCT
2551 GCAGGCATAG GGCAGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT
2601 CTCCTACGGA CGCTACAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651 AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCAGCT GGGCGCACTG
2701 GCGGGTGTCA ACGTTCCGTT TGCCGCAACG GGAGATTTGA CGGTCTGAAG
2751 CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCAATTCGC GAAAAAGGCA
2801 GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCCGA
2851 CTCGCGGGTC TGAAGCTGTC GCACCCCTTG AGCGATAAAG CCGTCTGTCT
2901 TGCAACGGCG GCGGTGGAAC GCGACCTGAA CCGACGCGAC TACACGPTAA
2951 CCGGCGGCTT TACCGGCGCG ACTGCAGCAA CCGCAAGAC GGGGGCACGC
3001 AATATGCCCG ACACCCGTCT GGTGCGCGCG CTGGGCGCGG ATGTCGAATT
3051 CCGCAACGGC TGGAACGGCT TGGCACGTTA CAGCTACGCC GGTTCCAAAC
3101 AGTACGGCAA CCACAGCGGA CGAGTCGGCG TAGGCTACCG GTTCTCTGAC
3151 GGTGGCGGAG GCACTGGATC CTCAGATTTG GCAAACGATT CTTTTATCCG
3201 GCAGGTTCTC GACCGTCAGC ATTTCGAACC CCGCGGAAA TACCACCTAT
3251 TCGGCAGCAG GGGGGAACCT GCGGAGCGCA GCGGCATAT CCGATTGGGA
3301 AAAATACAAA GCCATCAGTT GGGCAACCTG ATGATTCAAC AGGCGGCCAT
3351 TAAAGGAAAT ATCGGCTACA TTGTCCGCTT TTCCGATCAC GGGCACGAAG
3401 TCCATTCCCC CTTGACAAAC CATGCCCTAC ATTCCGATTG TGATGAAGCC
3451 GGTAGTCCCG TTGACGGATT TAGCCCTTAC CGCATCCATT GGGACGGATA
3501 CGAACACCAT CCCGCCGACG GCTATGACGG GCCACAGGGC GCGGCTATC
3551 CCGCTCCCAA AGGCGCGAGG GATATATACA GCTACGACAT AAAAGGCGTT
3601 GCCCAAAATA TCCGCTCAA CCGACCGAC AACCGCAGCA CCGGACAACG
3651 GCTTGGCGAC CGTTTCACA ATGCCGCTAG TATGCTGACG CAAGGAGTAG
3701 GCGACGGATT CAAACGCGCC ACCCGATACA GCGCGAGCT GGACAGATCG
3751 GGCAATGCCG CCGAAGCCTT CAACGGCACT GCAGATATCG TTAACAAACAT
3801 CATCGGCGCG GCAGGAGAAA TTGTCGGCGC AGGCGATGCC GTGCAGGCA
3851 TAAGCGAAGG CTCAAACATT GCTGTCATGC ACGGCTTGGG TCTGCTTTCC
3901 ACCGAAACAA AGATGGCGCG CATCAACGAT TTGGCAGATA TGGCGCAACT
3951 CAAAGACTAT GCCGACGAG CCATCCGCGA TTGGGCAGTC CAAAACCCCA
4001 ATGCCGCACA AGGCATAGAA GCGGTCAGCA ATATCTTTAT GGCAGCCATC
4051 CCCATCAAA GATTGGAGC TGTTCGGGGA AAATACGGCT TGGGCGGCAT
4101 CACGGCACAT CCTATCAAGC GGTGCGAGAT GGGCGCGATC GCATTGCCGA
4151 AAGGGAAATC CCGGCTCAGC GACAATTTTG CCGATCGGGC ATACGCCAAA
4201 TACCGTCCC CTTACCATTC CCGAAATATC CGTTCAAAC TGGAGCAGCG
4251 TTACGGCAA GAAAACATCA CCTCCTCAAC CGTGGCGCGG TCAAACGGCA
4301 AAAATGTCAA ACTGGCAGAC CAACGCCACC CGAAGACAGG CGTACCGTTT
4351 GACGGTAAAG GGTTCGGA TTTTGAGAAG CACGTGAAAT ATGATACGCT
4401 CGAGCACCAC CACCACCACC ACTGA

```

```

1   MTSAPDFNAG GTGIGSNSRA TTAASAASVY AGIKNEMCKD RSMLCAGRDD
51  VAVTDRDAKI NAPPFNLHTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV
101 GIVDTGESVG SISFPFLYGR KEHGYNENYK NYTAYMRKEA PEDGGGKDIE
151 ASFDDEAVIE TEAKPTDIRH VKEIGHIDLK SHIIGGRSVD GRPAGGIAPD
201 ATLHIMTND ETEKEMMVAA IRNAWVKLGE RGVRIVMNSF GTTSRAGTAD
251 LFQIANSEBQ YRQALLDYSG GDRKDEGIRL MQQSDYGNLS YHIRKNMLF
301 IFSTGNDAAQ QPNTYALLPF YEKDAQKGII TVAGVDRSGE KFKREMYGEP
351 GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG
401 TAALLLQKYP WMSNDNLRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG
451 PASFPFGDFT ADTKGTSDIA YSFRNDISGT GGLIKKGGSQ LQLHGNNTYT
501 GKTIIEGGSF VLVGNKSDM RVETKGALTY NGAASGGSIN SDGIVYLADT
551 DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTATIG GKLYMSARGK
601 GAGYLNSTGR RVPFLSAAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE
651 GDTLSYYVER GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
701 SSATPETVET AAADRTMPG IRPYGATFRA AAAYQHANA DGVRIFNSLA
751 ATVYADSTAA HADMQRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG
801 VEGKMRGSTQ TVGLAAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF
851 AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHABGSV NGTLMQLGAL
901 GGVNVFPAAT GDLTVEGGLR YDILLQDAFA EKGSLALWSG NSLTGGLTVG
951 LAGLKLSQPL SDKAVLFATA GVERDLNGRD YTVTGFTGA TAATGKTGAR
1001 NMPHTRLVAG LGADVEFGNG WNGIARYSYA GSKQYGNHSG RVGVGYRFLD
1051 GGGGTGSSDL ANDSFIRQVL DRQEFEPDGE YHLFGSRGEL AERSGHIGLG
1101 KIQSHQLGNL MIQQAALKGN IGYIVRFSDE GHEVHSPPDN HASHSDSDEA
1151 GSPVDGFSLY RIHMDGYEHH PADGYDGPQG GGYPAPKGAR DIYSYDIKGV
1201 AQNIRLNLFD NRSTGQRLAD RFHNAGSMLT QGVGDGPKRA TRYSPELDRS
1251 GNAABAFNGT ADIVKNIIGA AGEIVGAGDA VQGISSEGSNI AVMHGLGLLS
1301 TENKMARIND LADMAQLKDY AAAAIRDNAY QNFNAAQGIE AVSNIFMAAI
1351 PIKIGIAGVRG KYGLGGITAH PIKRSQMGAI ALPRGKSAVS DNFADAAYAK
1401 YPSFYHSENI RSNLEQRYGK ENITSSTVPP SNGKNVKLAD QRHFKTGVPF
1451 DGKGFPNFEK HVKYDTLEHH HHHH*

```

AG983-741

```

1   ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
51  CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
101 AGAAGCAAAAT GTGCAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC
151 GTTGCCTGTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCAGATCT
201 GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTGTATCA
251 ACCTCAAACC TGCAATGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
301 GGTATCGTCG ACACAGGCGA ATCCGTGCGC AGCATATCCT TTCCGGAAT
351 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
401 CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
451 GCTTCTTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
501 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
551 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT
601 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACAGAAATGAT
651 GGTTCAGACC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
701 GCATCGTCAA TAACAGTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
751 CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
801 CTATTCGGCG GGTGATAAAA CAGACGAGGG TATCCGCCTG ATGCAACAGA
851 CCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAAA CATGCTTTTC
901 ATCTTTTCGA CAGGCAATGA CGCACAGCT CAGCCCAACA CATATGCCCT
951 ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001 GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTCGGGAA TTAAGTCCAT
1101 GTGGTGCCTG TCGGCACCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
1151 ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC
1201 ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
1251 GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG
1301 ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CCGGTAAGGC CATGAACGGA
1351 CCCGCGTCTT TCCGTTTCGG CGACTTTACC GCCGATACGA AAGGTATATC
1401 CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GCGCGCTGA
1451 TCAAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501 GGCAAAACCA TTATCGAAGG CGGTTCCGTC GTGTGTACG GCAACACAA
1551 ATCGGATATG CGGCTCGAAA CCAAGGTGC GCTGATTTAT AACGGGCGG
1601 CATCCGGCGG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC
1651 GACCAATCCG GCGCAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT

```

```

1701 GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAAC TG GAAAGTGG
1751 ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG
1801 GGGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCCT TCCTGAGTGC
1851 CGCCAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
1901 GCGGCCTGCT GGCTTCCCTC GACAGCGTCG AAAAAACAGC GGGCAGTGAA
1951 GGCACACGCG TGTCCTATTA TGTCCGTCGC GGCAATGCGG CACGGACTGC
2001 TTCGGCAGCG GCACATTCCG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC
2051 AGGGCGGCGC CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101 TCATCCGCAA CACCCGAGAC GGTGAAACT GCGGCAGCCG ACCGCACAGA
2151 TATGCCGGGC ATCCGCCCCT ACGGCGCAAC TTTCGGGCA GCGGCAGCCG
2201 TACAGCATGC GAATGCCGCC GACGGGTGAC GCATCTTCAA CAGTCTCGCC
2251 GCTACCGTCT ATCCGACAG TACCGCCGCC CATGCCGATA TGCAGGGACG
2301 CCGCCTGAAA GCCGTATCGG ACGGGTTGGA CCACAACGGC ACGGGTCTGC
2351 GCGTTCATCGC GCAAAACCAA CAGGACCGTG GAACGTGGGA ACAGGGCGGT
2401 GTTGAAGGCA AAATGCGCGG CAGTACCCAA ACCGTCGGCA TTGCCGCGAA
2451 AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGCGACA
2501 CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTTT
2551 GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCCTGTT
2601 CTCCTACCGA CGCTACAAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651 AACATGCGGA AGGCAGCGTC AACGGCACCG TGATGCAGCT GGGCGCACTG
2701 GCGCGGTGTA ACGTTCGTTT TGCCGCAACG GGAGATTGGA CGGTGGAAGG
2751 CGGTCTCGCG TACGACCTGC TCAAAACAGG TGCATTCGCC GAAAAAGGCA
2801 GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGGTCGGA
2851 CTCGCGGGTC TGAAGCTGTC GCAACCCCTG AGCGATAAAG CCGTCCGTGT
2901 TGCAACGGCG GGCCTGGAAC GCGACCTGAA CCGACGCGAC TACACGGTAA
2951 CGGGCGGCTT TACCGCGCGC ACTGCAGCAA CCGCAAGAC GGGGCGACGC
3001 AATATGCCGT ACACCCGTCT GGTGCGCGCG CTGGGCGCGG ATGTCGAATT
3051 CGGCAACGGC TGGAACGGCT TGGCACGTTA CAGCTACGCC GGTTCCAAAC
3101 AGTACGGCAA CCACAGCGGA CGAGTCGCGC TAGGCTACCG GTTCCTCGAG
3151 GGATCCGGAG GGGGTGGTGT CGCCGCGGAC ATCGGTGCGG GGCTTGCCGA
3201 TGCATAACC GCACCGCTCG ACCATAAGA CAAAGGTTTG CAGTCTTTGA
3251 CGCTGGATCA GTCCGTCAGG AAAAAACGGA AACTGAAGCT GCGGCGACAA
3301 GGTGCGGAAA AAACCTATGG AAACGGTGAC AGCCTCAATA CCGGCAAATT
3351 GAAGAACGAC AAGGTCAGCC GTTTCGACTT TATCCGCCAA ATCGAAGTGG
3401 ACGGGCAGCT CATACCTTG GAGAGTGGAG AGTTCCAAGT ATACAAACAA
3451 AGCCATTCCG CCTTAACCGC CTTTCAGACC GAGCAAATAC AAGATTGCGA
3501 GCATTCCGGG AAGATGGTTG CGAAACGCCA GTTCAGAATC GCGACATAG
3551 CGGGCGAACA TACATCTTTT GACAAGCTTC CCGAAGGCGG CAGGGCGACA
3601 TATCGCGGGA CGGCGTTCGG TTCAGACGAT GCCGCGGGA AACTGACCTA
3651 CACCATAGAT TTCGCGCCA AGCAGGAAA CCGCAAATC GAACATTGGA
3701 AATCGCCAGA ACTCAATGTC GACCTGGCCG CCGCCGATAT CAAGCCGAT
3751 GGAAACGCC ATGCCGTCAT CAGCGGTTCC GTCTTTTACA ACCAAGCCGA
3801 GAAAGGCAGT TACTCCCTCG GTATCTTTGG CGGAAAAGCC CAGGAAAGTG
3851 CCGGCAGCGC GGAAGTGAAA ACCGTAAACG GCATACGCCA TATCGGCCTT
3901 GCCGCCAAGC AACTCGAGCA CCACCACCAC CACCACTGA

```

```

1 MTSAPDFNAG GTGIGSNSRA TTAksAAVSy AGIKNEMCKD RSMLCAGRDD
51 VAVTDRDAKI NAPPNLTG DFPNPNDAYK NLINLKPAIE AGYTGRGVEV
101 GIVDTGESVG SISFPelyGR KKHGYNENYK NYTAYMRKEA PEDGGGKDIE
151 ASFDDEAVIE TEAKPTDIRH VKEIGHIDLv SHIIGGRSVD GRPAGGIAPD
201 ATLHIMNTND ETKNEMVAA IRNANVLGE RGVRIvMNSF GTTSRAGTAD
251 LFQIANSEBQ YRQALLDYSG GDKTDEGIRL MQQSDYGNLS YHIRNKNMLF
301 IPSTGNDAQA QPNTYALLPF YEKDAQKGIi TVAGVDRSGE KFKREMYGEP
351 GTEPLEYGSN HCGITAMWCL SAPYEASVRf TRTNPIQIAG TSFSAPIVTG
401 TAALLLQKYP WMSNDNLRTT LLTTAQDIGA VGVDSEKFGWg LLDAGKAMNG
451 PASFPFGDFT ADTKGTSdia YSFRNDISGT GGLIKKGGsQ LQLHGNTYTT
501 GKTIEGGSL VLYGNKSDM RVETKGALIY NGAASGGSLN SDGIVYLADT
551 DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDGTAIIG GKLYMSARGK
601 GAGYLNSTGR RVFFLBAAKI QDYSFFTNi ETDGGLLASL DSVEKTAGSE
651 GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
701 SSATPETVET AAADRTMPG IRPYGATFRA AAaVQHANAa DOVRIFNSLA
751 ATVYADSTAA HADMQGRRLK AVSDGLDENG TGLRVIAQTQ QDGGTMEQGG
801 VEGKMRGSTQ TVGLAAKTGE NTTAAATLGM GRSTWSENsA NAKTDSISLF
851 AGIRHDAGDI GYLKGLFSYG RYKNSISRSr GADEHABGSV NOTLMQLGAL
901 GGVNVFFAAT GDLTVEGGLR YDLLKQDAPa EKGSALGWSG NSLTBGTLVG
951 LAGLKLsqPL SDKAVLFATA GVERDLNGRD YTVTGGFTGa TAATGKTGAR
1001 NMPHTRLVAG LGADVEFGNG WNGLARYSYa GSKQYGNHSG RVGVGYRFLB

```

```

1051 GSGGGGVAAD IGAGLADALT APLDHRDKGL QSLTLDSQSVR XNEKLKLAAG
1101 GAECTYGNND SLNTGKLKND KVSRLFDFIRQ IEVDGQLITL ESGEFQVYKQ
1151 SHSALTAFQT EQIQDSEHSG KMKAKRQFRI GDIAGEHTSF DKLPEGGRAT
1201 YRGTAFGSDD AGGKLYTTID FAAKQGNKRI EHLKSPELNV DLAAADIKPD
1251 GKRHAVISGS VLYNQAEKGS YSLGIFGKA QEVAGSAEVK TVNGIRHIGL
1301 AAKQLEHHHH HH*

```

AG983-961

```

1 ATGACTTCTG CGCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
51 CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
101 AGAAGCAAT GTGCAAGAC AGAAGCATGC TCTGTGCCGG TCGGATGAC
151 GTTGCCGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT
201 GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTGTATCA
251 ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GGTAGAGGTA
301 GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCCGAACT
351 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
401 CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
451 GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
501 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
551 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TCGCGCCGAT
601 CGCAGCTACT ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
651 GGTTCGACCC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
701 GCATCGTCAA TAACAGTTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
751 CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCTCG ATGCACAGA
851 CGGATTACGG CAACCTGTTC TACCACATCC GTAATAAAAA CATGCTTTTC
901 ATCTTTTCGA CAGGCAATGA CGCACAAAGC CAGCCCAACA CAATGCCCC
951 ATTGCCATTT TATGAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001 GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051 GGTACAGAAC CGCTGAGTA TGGCTCAAAC CATTCGCGAA TTACTGCCAT
1101 GTGGTGCTCG TCGGCACCTT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
1151 ACCCGATTCA AATTGCGGGA ACATCCTTTT CCGCACCCAT CGTAACCGCG
1201 ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGGATGAGCA ACGACAACCT
1251 GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCCGCGTGG
1301 ACAGCAAGTT CGCTGGGGA CTGCTGGATG CGGTAAGGC CATGAAACGA
1351 CCGCGTCTCT TTCCGTTCCG CGACTTTACC GCCGATACGA AAGGTACATC
1401 CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GCGCGCTGA
1451 TCAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501 GGCAAAACCA TTATCGAAGG CGGTTCGCTG GTGTGTACG GCAACAACAA
1551 ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTTAT AACGGGGCGG
1601 CATCCGGCGC CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC
1651 GACCAATCCG GCGCAACGA AACCGTACAC ATCAAAGGCA GTCTGCAGCT
1701 GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG
1751 ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGGCAAG
1801 GGGCAGGCT ATCTCAACAG TACCGGACGA CGTGTTCCTT TCCTGAOTGC
1851 CGCCAAAATC GGGCAGGATT ATTCTTTCTT CACAAACATC GAAACCGACG
1901 CGCGCCTGCT GCGTTCCTTC GACAGCGTCG AAAAACAGC GGGCAGTGAA
1951 GCGACACGC TGTCCTATTA TGTCCGTCGC GGCAATGCGG CACGGACTGC
2001 TTCGGCAGCG GCACATTCGG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC
2051 AGGGCGGCGA CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101 TCATCCGCAA CACCCGAGAC GGTGAAACT CGGCAGCCG ACCGCACAGA
2151 TATGCCGGGC ATCCGCCCTT ACGGCGCAAC TTTCGCGCA GCGGCGCCG
2201 TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
2251 GCTACCGTCT ATGCCGACAG TACCGCCGCC CATGCCGATA TGCAGGACG
2301 CCGCTGAAA GCCGTATCGG ACGGGTTGGA CCACAACGGC ACGGPTCTGC
2351 GCGTCATCGC GCAAACCCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
2401 GTTGAAGGCA AAATCGCGCG CAGTACCCAA ACCGTGCGCA TTGCCGCGAA
2451 AACC GGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGACGA
2501 CATGGAGCGA AAACAGTGCA AATGCAAAA CCGACAGCAT TAGTCTGTTT
2551 GCAGGCATAC GGCACGATGC GGGCGATATC GGCATATCTA AAGGCCGTGT
2601 CTCCTACGGA CGCTACAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651 AACATGCGGA AGGCGCGTC AACGGCACG TGATGCAGCT GGGCGCACTG
2701 GCGCGTGTCA ACGTTCGTT TGCCGCAACG GGAGATTGGA CGGTGCAAGG
2751 CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCAATCGCC GAAAAAGGCA
2801 GTCTTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC OCTGGTCCGA
2851 CTCGCGGGTC TGAAGCTGTC GCAACCCTTG AGCGATAAAG CCGTCTGTT

```

```

2901 TGCAACGGCG GGCCTGGAAC GCGACCTGAA CGGACGCGAC TACACGGTAA
2951 CGGGCGGCTT TACCGCGCGG ACTGCAGCAA CCGGCAAGAC GGGGGCAGCG
3001 AATATGCCGC ACACCCGTCT GGTTCGCCGC CTGGGCGCGG ATGTCGAATT
3051 CGCAACGGCG TGGAACGGCT TGGCACGTTA CAGCTACGCC GGTTCCAAAC
3101 AGTACGGCAA CCACAGCGGA CGAGTCGGCG TAGGCTACCG GTTCCTCGAG
3151 GGTGGCGGAG GCACTGGATC CGCCACAAAC GACGACGATG TAAAAAAGC
3201 TGCCACTGTG GCCATTGCTG CTGCCTACAA CAATGGCCAA GAAATCAACG
3251 GTTTCAAAGC TGGAGAGACC ATCTACGACA TTGATGAAGA CGGCACAATT
3301 ACCAAAAAAG ACGCAACTGC AGCCGATGTT GAAGCCGACG ACTTTAAAGG
3351 TCTGGGTCTG AAAAAAGTCG TGACTAACCT GACCAAAACC GTCAATGAAA
3401 ACAAACAAAA CGTCGATGCC AAAGTAAAG CTGCAGAATC TGAATAGAA
3451 AAGTTAACAA CCAAGTTAGC AGACACTGAT GCCGCTTTAG CAGATACTGA
3501 TGCCGCTCTG GATGCAACCA CCAACGCCCT GAATAAATTG GGAGAAAATA
3551 TAACGACATT TGCTGAAGAG ACTAAGACAA ATATCGTAAA AATTGATGAA
3601 AAATTAGAAG CCGTGGCTGA TACCGTCGAC AAGCATGCCG AAGCATTCAA
3651 CGATATCGCC GATTCAATGG ATGAAACCAA CACTAAGGCA GACGAAGCCG
3701 TCAAAACCGC CAATGAAGCC AAACAGACGG CCGAAGAAAC CAAACAAAC
3751 GTCGATGCCA AAGTAAAGC TGCAGAAACT GCAGCAGGCA AAGCCGAAGC
3801 TGCCGCTGGC ACAGCTAATA CTGCAGCCGA CAAGGCCGAA GCTGTCGCTG
3851 CAAAAGTTAC CGACATCAAA GCTGATATCG CTACGAACAA AGATAATATT
3901 GCTAAAAAAG CAAACAGTGC CGACGTGTAC ACCAGAGAAG AGTCTGACAG
3951 CAAATTTGTC AGAATTGATG GTCTGAACGC TACTACCGAA AAATTGGACA
4001 CACGCTTGGC TTCTGCTGAA AAATCCATTG CCGATCACGA TACTCGCCTG
4051 AACGTTTGG ATAAACAGT GTCAGACCTG CGCAAGAAA CCCGCCAAGG
4101 CCTTGCAGAA CAAGCCGCGC TCTCCGGTCT GTTCCAACCT TACAACGTGG
4151 GTCGGTTCAA TGTAACGGCT CGAGTCGGCG GCTACAAATC CGAATCGGCA
4201 GTCGCCATCG GTACCGGCTT CCGCTTTACC GAAAACTTTG CCGCCAAAGC
4251 AGGCGTGGCA GTCGGCACTT CGTCCGGTTC TTCCGCAGCG TACCATGTCTG
4301 GCGTCAATTA CGAGTGGCTC GAGCACCACC ACCACCACCA CTGA

```

```

1 MTSAPDFNAG GTGIGNSNSRA TTKSAAVSY AGIKNEMCKD RSMLCAGRDD
51 VAVTDRDAKI NAPPFNLETG DFFNPNDAKY NLINLKPAIE AGYTGGRVVE
101 GIVDTGESVG SISFPFLYGR KEHGYNENYK NYTAYMRKEA PEDGGGKDIE
151 ASFDDEAVIE TEAKPTDIRH VKEIGHIDLK SHIIGGRSVD GRPAGGIAPD
201 ATLHIMNTND ETKNEMVAA IRNANVKLGE RGVRIVMNSF OTTSRAGTAD
251 LFQIANSEBQ YRQALLDYSG GDRTEDEGRL MQQSDYGNLS YHIRNKMMLF
301 IFSTGNDAQA QPNTYALLPF YEKDAQEGII TVAGVDRSGE KFKREMYGEP
351 GTEPLEYGSN HCGITAMWCL SAPYEASVRF TRTNPIQIAG TSFSAPIVTG
401 TAALLLQKYP WMSNDMLRTT LITTAQDIGA VGVDSKFGWG LLDAGKAMNG
451 PASFFFGDFT ADTKGTSIDIA YSFRNDISGT GGLIKKGSQ LQLHGNFTYT
501 GKTIEGGSL VLYGNMSDM RVETKGALII NGAASSGSLN SDGIVYLADT
551 DQSGANETVH IKGSLQLDGK GTLYTRLGKL LKVDQTAIIG GKLYMSARGK
601 GAGYINSTGR RVFFLSAAKI GQDYSFFTNI ETDGGLLASL DSVEKTAGSE
651 GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSNLE NLMVELDASE
701 SSATPFTVET AAADRTMPG IRPYGATFRA AAAVQHANA ADOVRIFNSLA
751 ATVYADSTAA HADMQRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG
801 VEGKMRGSTQ TVGLAAKTGE NTAATAATLGM GRSTWSENBA NAKTDSISLF
851 AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADEHAEGSV NGTLMQLGAL
901 GGVNVFFAAT GDLTVEGGLR YDLLKQDAFA EKGSAWGWSG NSLTBOTLVG
951 LAGLKLSQLP SDKAVLFATA GVERDLNGRD YTVTGFTGA TAATGKTGAR
1001 NMPKTRLVAG LGADVFPNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLB
1051 GGGGTGSATN DDDVKEAATV AIAAAYNNGQ EINGFKAGET IYDIDEDGTI
1101 TKKDATAADV EADDFKGLGL KKVVTNLTKT VNENKQNVDA KVKAESIE
1151 KLTTLADTD AALADTDAAL DATTNALNKL GENITTFABE TKTNIKIDE
1201 KLEAVADTV DHAFAFNDIA DSLDETNTKA DEAVKTANEA KQTAETKQN
1251 VDAKVKAAET AAGKAEEAAG TANTAADKAE AVAAKVTDIK ADIATNKDNI
1301 AKKANSADV TRESDSKFV RIDGLNATTE KLDTRLASAE KSLADEHTRL
1351 NGLDKTVSDL RKETRQGLAE QALSGLFQP YNVGRFNVTA AVGGYKESBA
1401 VAIGTGFRFT ENFAAKAGVA VOTSSGSSAA YHVGNYEWL EHHHHH*

```

AG983-961c

```

1 ATGACTTCTG CCCCCGACTT CAATGCAGGC GGTACCGGTA TCGGCAGCAA
51 CAGCAGAGCA ACAACAGCGA AATCAGCAGC AGTATCTTAC GCCGGTATCA
101 AGAACGAAAT GTGCAAGAC AGAAGCATGC TCTGTGCCGG TCGGGATGAC
151 GTTCGGGTTA CAGACAGGGA TGCCAAAATC AATGCCCCCC CCCCGAATCT
201 GCATACCGGA GACTTTCCAA ACCCAAATGA CGCATACAAG AATTGTATCA

```

```

251 ACCTCAAACC TGCAATTGAA GCAGGCTATA CAGGACGCGG GOTAGAGGTA
301 GGTATCGTCG ACACAGGCGA ATCCGTCGGC AGCATATCCT TTCCGAACT
351 GTATGGCAGA AAAGAACACG GCTATAACGA AAATTACAAA AACTATACGG
401 CGTATATGCG GAAGGAAGCG CCTGAAGACG GAGGCGGTAA AGACATTGAA
451 GCTTCTTTCG ACGATGAGGC CGTTATAGAG ACTGAAGCAA AGCCGACGGA
501 TATCCGCCAC GTAAAAGAAA TCGGACACAT CGATTTGGTC TCCCATATTA
551 TTGGCGGGCG TTCCGTGGAC GGCAGACCTG CAGGCGGTAT TGCGCCCGAT
601 GCGACGCTAC ACATAATGAA TACGAATGAT GAAACCAAGA ACGAAATGAT
651 GGTTCAGACC ATCCGCAATG CATGGGTCAA GCTGGGCGAA CGTGGCGTGC
701 GCATCGTCAA TAACAGTTT GGAACAACAT CGAGGGCAGG CACTGCCGAC
751 CTTTTCCAAA TAGCCAATTC GGAGGAGCAG TACCGCCAAG CGTTGCTCGA
801 CTATTCCGGC GGTGATAAAA CAGACGAGGG TATCCGCTG ATGCAACAGA
851 GCGATTACGG CAACCTGTCC TACCACATCC GTAATAAAAA CATGCTTTTC
901 ATCTTTTCGA CAGGCAATGA CGCACAGCT CAGCCCAACA CATATGCCCT
951 ATTGCCATTT TATGAAAAAG ACGCTCAAAA AGGCATTATC ACAGTCGCAG
1001 GCGTAGACCG CAGTGGAGAA AAGTTCAAAC GGGAAATGTA TGGAGAACCG
1051 GGTACAGAAC CGCTTGAGTA TGGCTCCAAC CATTCGCGAA TTAAGTCCAT
1101 GTGGTGCCCTG TCGGCACCCCT ATGAAGCAAG CGTCCGTTTC ACCCGTACAA
1151 ACCCGATTCA AATTGCCGGA ACATCCTTTT CCGCACCCAT CGTAACCGGC
1201 ACGGCGGCTC TGCTGCTGCA GAAATACCCG TGATGAGCA ACGACAACTT
1251 GCGTACCACG TTGCTGACGA CGGCTCAGGA CATCGGTGCA GTCGGCGTGG
1301 ACAGCAAGTT CGGCTGGGGA CTGCTGGATG CGGGTAAGGC CATGAACGGA
1351 CCCGCGTCTT TCCGTTCCG CGACTTTACC GCCGATACGA AAGGTACATC
1401 CGATATTGCC TACTCCTTCC GTAACGACAT TTCAGGCACG GCGCGCTGA
1451 TCAAAAAGG CGGCAGCCAA CTGCAACTGC ACGGCAACAA CACCTATACG
1501 GGCAAAACCA TTATCGAAGG CGGTTCCGTC GTGTTGTACG GCAACAACAA
1551 ATCGGATATG CGCGTCGAAA CCAAAGGTGC GCTGATTAT AACGGGCGCG
1601 CATCCGCGCG CAGCCTGAAC AGCGACGGCA TTGTCTATCT GGCAGATACC
1651 GACCAATCCG GCGCAACGA AACCCTACAC ATCAAAGGCA GTCTGCAGCT
1701 GGACGGCAAA GGTACGCTGT ACACACGTTT GGGCAAACTG CTGAAAGTGG
1751 ACGGTACGGC GATTATCGGC GGCAAGCTGT ACATGTCGGC ACGCGCAAG
1801 GGGCGAGGCT ATCTCAACAG TACCGGACGA CGTGTCCCT TCCTGAGTGC
1851 CGCAAAATC GGGCAGGAT ATCTTTTCTT CACAAACATC GAAACCGACG
1901 GCGGCTGCT GGGTTCCCTC GACAGCGTCG AAAAACAGC GGGCAOTGAA
1951 GGGGACACGC TGTCTTATTA TGTCCGTCGC GGCATGCGG CACGGACTGC
2001 TTCGGCAGCG GCACATTCGG CGCCCGCCGG TCTGAAACAC GCCGTAGAAC
2051 AGGGCGGCGC CAATCTGGAA AACCTGATGG TCGAACTGGA TGCCTCCGAA
2101 TCATCCGCAA CACCCGAGAC GGTGAAACT GCGGCAGCCG ACCGACAGAA
2151 TATGCCGGGC ATCCGCCCTT ACGGCGCAAC TTTCCGCGCA GCGGCGCCG
2201 TACAGCATGC GAATGCCGCC GACGGTGTAC GCATCTTCAA CAGTCTCGCC
2251 CGTACCGTCT ATGCGGACAG TACCGCCGCC CATGCCGATA TGCAGGGACG
2301 GCGCTTGAAA GCCGTATCGG ACGGGTTGGA CCACAAACGC ACGGGTCTGC
2351 GCGTCAATCG GCAAAACCAA CAGGACGGTG GAACGTGGGA ACAGGGCGGT
2401 GTTGAAGCCA AAATGCGCGG CAGTACCCAA ACCGTGGGCA TTGCCGCGAA
2451 AACCGGCGAA AATACGACAG CAGCCGCCAC ACTGGGCATG GGACGCAGCA
2501 CATGGAGCGA AAACAGTGCA AATGCAAAAA CCGACAGCAT TAGTCTGTTT
2551 GCAGGCATAC GGCACGATGC GGGCGATATC GGCTATCTCA AAGGCTGTTT
2601 CTCCTACGGA CGCTACAAAA ACAGCATCAG CCGCAGCACC GGTGCGGACG
2651 AACATGCGGA AGGCAGCGTC AACGGCACGC TGATGCAGCT GGGCGCACTG
2701 GGGCGTGTCA ACGTTCCGTT TGCCGCAACG GGAGATTGTA CGGTGGAAGG
2751 CGGTCTGCGC TACGACCTGC TCAAACAGGA TGCATTGCCC GAAAAAGGCA
2801 GTGCTTTGGG CTGGAGCGGC AACAGCCTCA CTGAAGGCAC GCTGTCGGA
2851 CTGCGGGGTC TGAAGCTGTC GCAACCCTTG AGCGATAAAG CCGTCTGTTT
2901 TGCAACGGCG GCGTGGAAC GCGACCTGAA CGGACCGGAC TACACGOTAA
2951 CGGGCGGCTT TACCGGCGCG ACTGCAGCAA CCGGCAAGAC GGGGGCACGC
3001 AATATGCCGC ACACCCGTCT GGTGCGCGC CTGGGCGCGG ATGTCGAATT
3051 CGGCAACGGC TGGAACGGCT TGGCAGTTA CAGCTACGCC GGTTCCAAAC
3101 AGTACGGCAA CCACAGCGGA CGAGTCGGCG TAGGCTACCG GTTCTCGAG
3151 GGTGGCGGAG GCACTGGATC CGCCACAAAC GACGACGATG TTAATAAAGC
3201 TGCCACTGTG GCCATTGCTG CTGCCTACAA CAATGCCCAA GAAATCAACG
3251 GTTTCAAAGC TGGAGAGACC ATCTACGACA TTGATGAAGA CGGCACAATT
3301 ACCAAAAAG ACGCCAATGC AGCCGATGTT GAAGCGGACG ACTTTAAAGG
3351 TCTGGGTCTG AAAAAAGTCG TGAATAACCT GACCAAAACC GTCAATGAAA
3401 ACAAACAAA CGTCGATGCC AAAAGTAAAG CTGCAGAAAT TGAAATAGAA
3451 AAGTTAACAA CCAAGTTAGC AGACACTGAT GCCGCTTAG CAGATACTGA
3501 TGCCGCTCTG GATGCAACCA CCAACGCCTT GAATAAATTG GGAGAAAATA
3551 TAACGACATT TGCTGAAGAG ACTAAGACAA ATATCGTAAA AATTGATGAA

```

```

3601 AAATTAGAAG CCGTGGCTGA TACCGTCGAC AAGCATGCCG AAGCATTCAA
3651 CGATATCGCC GATTCATTGG ATGAAACCAA CACTAAGCCA GACGAAGCCG
3701 TCAAAACCGC CAATGAAGCC AAACAGACGG CCGAAGAAAC CAAACAAAAC
3751 GTCGATGCCA AAGTAAAAGC TGCAGAAACT GCAGCAGGCA AAGCCGAAAG
3801 TGCCGCTGGC ACAGCTAATA CTGCAGCCGA CAAGGCCGAA GCTGTGCGCTG
3851 CAAAAGTTAC CGACATCAAA GCTGATATCG CTACGAACAA AGATAATATT
3901 GCTAAAAAAG CAAACAGTGC CGACGTGTAC ACCAGAGAAG AGTCTGACAG
3951 CAAATTTGTC AGAATTGATG GTCTGAACGC TACTACCGAA AAATTGGACA
4001 CACGCTTGGC TTCTGCTGAA AAATCCATTG CCGATCACGA TACTCGCCTG
4051 AACGGTTTGG ATAAACAGT GTGACAGCTG CGCAAAGAAA CCCGCCAAGG
4101 CCTTGCAGAA CAAGCCGCGC TCTCCGGTCT GTTCCAACCT TACAACGTGG
4151 GTCTCGAGCA CCACCACCAC CACCACTGA

```

```

1 MTSAPDFNAG GTGIGSNSRA TTAASAASVY AGIKNEMCKD RSMLCAGRDD
51 VAVTDRDAKI NAPPFNLHTG DFPNPNDAYK NLINLKPATB AGYTGRGVFV
101 GIVDTGHSVG SISFFELYGR KEHGYNENYK NYTAYMKRA PEDGGGIDIE
151 ASFDDEAVIE TEAKPTDIRH VKEIGHIDLK SHIIGORSVD GRPAGGIAPD
201 ATLEHMTFND ETKNEMVAA IRNAWVKLGE RGVRLVNNST GTTSHAGTAD
251 LFQIANSEEQ YRQALLDYSG GDKTDEGIRL MQQSDYGNLS YHIRKQMLF
301 IFSTGNDAAQ QPNTYALLPF YEKDAQKGI TVAGVDRSGR KFKREMYGEP
351 GTEPLEYOSN HCGITAMMCL SAPYEASVRF TRTNPIQLAG TSFSAPIVIG
401 TAAALLQKYP WMSNDMLRTT LLTTAQDIGA VGVDSKFGWG LLDAGKAMNG
451 PASFFPGDFT ADTEGTSDLA YSFRNDISGT GGLIKKGGSQ LQLHGNFTYT
501 GKTIEBGGSL VLYGNKSDM RVETKGALTY NGAASGGSLN SDGIVYLADT
551 DQSGANETVE IKGSLQLDGF GTLYTRLGKL LKVDGTALIG GKLYMSARGK
601 GAGYLNSTGR RVPFLSAKI GDYSFFTHI ETDGGLLASL DSVETAGSSE
651 GDTLSYYVRR GNAARTASAA AHSAPAGLKH AVEQGGSMLE NLMVELDASE
701 SSATFETVET AAADRTMPG IRPYGATFRA AAAVQEAHAA DGVRLFNLSA
751 ATVIADSTAA HADMQRRLK AVSDGLDHNG TGLRVIAQTQ QDGGTWEQGG
801 VEGKMRGSTQ TVGLIAKTGE NTTAAATLGM GRSTWSENSA NAKTDSISLF
851 AGIRHDAGDI GYLKGLFSYG RYKNSISRST GADSHAGSV NGTLMQLGAL
901 GGVNVFFAAT GDLTVEGGLR YDLLKQDAFA EKGSAWGWSG NSLTBGTLVG
951 LAGLKLSQLP SDAVLFAFA GVERDLNGRD YTVTGGFTGA TAATGKTGAR
1001 NMPHTRLVAG LGADVEFGNG WNGLARYSYA GSKQYGNHSG RVGVGYRFLF
1051 GGGGTGSATN DDDVKAATV ALAAAYNNGQ EINGFKAGST IYDIDEDGTI
1101 TKKDATAADV EADDFEGLGL KKVVTNLTFT VNEKNQNVDA KVKAAESEIE
1151 KLTTKLADTD AALADTDAAL DATTNALNKL GENITTFAKE TKTNIVKIDE
1201 KLEAVADTV DHAFAFDIA DSLDETNTKA DEAVITANEA KQTAKETKQN
1251 VDAKVKAAT AAGKABAAAG TANTAADKAE AVAAKVTDLK ADIATNKDNI
1301 ARKANSADV TRESDSKPV RIDGLNATTE KLDTRLASAE KSLADHDTRL
1351 NGLDKTVSDL RKETRQGLAE QAALSGLFQP YNVGLEHHHH HH*

```

△G741 和杂交体

针对各种菌株(包括同源 2996 菌株),测定由应答△G741(His-融合)产生的杀菌滴定度:

	2996	MC58	NGH38	F6124	BZ133
△G741	512	131072	>2048	16384	>2048

5 可以看出,针对异源菌株 MC58,由△G741 引发的抗-杀菌滴定度特别高。

还将△G741 直接融合于蛋白质 961、961c、983 和 ORF46.1 的符合读框的上游:

△G741-961

```

1 ATGGTCGCCG CCGACATCGG TCGGGGCTT GCCGATGCAC TAACCGCACC
51 GCTCGACCAT AAAGACAANG GTTTCAGTC TTGACGCTG GATCAGTCCG
101 TCAAGAAAAA CGAGAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAACT
151 TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAAGT
201 CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA

```

```

251 CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
301 ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAGAT
351 GGTTCGGA AAA CGCCAGTTCA GAATCGGCGA CATAGCGGCG GAACATACAT
401 CTTTGTGACAA GCTTCCCGAA GGCGGCAGGG CGACATATCG CGGGACGGCG
451 TTCGGTTCAG ACGATGCCGG CGGAAACTG ACCTACACCA TAGATTTTCG
501 CGCCAAGCAG GGAACCGCA AAATCGAACA TTGAAATCG CCAGAACTCA
551 ATGTCGACCT GGCCGCCGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
601 GTCATCAGCG GTTCCGTCTT TTACAACCAA GCCGAGAAAG GCAGTTACTC
651 CCTCGGTATC TTTGGCGGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
701 TGA AAACCGT AAACGGCATA CGCCATATCG GCCTTGCCGC CAAGCAACTC
751 GAGGGTGGCG GAGGCACTGG ATCCGCCACA AACGACGACG ATGTTAAAAA
801 AGCTGCCACT GTGGCCATTG CTGCTGCTTA CAACAATGCG CAAGAAATCA
851 ACCGTTTCAA AGCTGGAGAG ACCATCTACG ACATTGATGA AGACGGCACA
901 ATTACCAAAA AAGACGCAAC TGCAGCCGAT GTTGAAGCCG ACGACTTTAA
951 AGGTCTGGGT CTGAAAAAAG TCGTGACTAA CCTGACCAA ACCGTCATG
1001 AAAACAAACA AACGTCGAT GCCAAAGTAA AAGCTGCAGA ATCTGAAATA
1051 GAAAAATTAA CAACCAAGTT AGCAGACACT GATGCCCGTT TAGCAGATAC
1101 TGATGCCCGT CTGGATGCAA CCACCAACGC CTTGAATAAA TTGGGAGAAA
1151 ATATAACGAC ATTGTCTGAA GAGACTAAGA CAAATATCGT AAAAATTGAT
1201 GAAAAATTAG AAGCCGTGGC TGATACCGTC GACAAGCATG CCGAAGCATT
1251 CAACGATATC GCCGATTCTT TGGATGAAAC CAACACTAAG GCAGACGAAG
1301 CCGTCAAAAC CGCCAATGAA GCCAAACAGA CGGCCGAAGA AACCAAAACA
1351 AACGTCGATG CCAAAGTAAA AGCTGCAGAA ACTGCAGCAG GCAAGCCGA
1401 AGCTGCCGCT GGCACAGCTA ATACTGCAGC CGACAAGGCC GAAGCTGTCTG
1451 CTGCAAAAGT TACCGACATC AAAGCTGATA TCGCTACGAA CAAGATAAT
1501 ATTGCTAAAA AAGCAAACAG TGCCGACGTG TACACCAGAG AAGAGTCTGA
1551 CAGCAAATTT GTGAGAATTG ATGGTCTGAA CGCTACTACC GAAAAATTGG
1601 ACACACGCTT GGTCTCTGCT GAAAAATCCA TTGCCGATCA CGATACTCGC
1651 CTGAACGGTT TGGATAAAAC AGTGTCAAGC CTGCGCAAAG AAACCGCCA
1701 AGGCCTTGCA GAACAAGCCG CGCTCTCCGG TCTGTTCCAA CCTTACAACG
1751 TGGGTCGGTT CAATGTAACG GCTGCAGTCG GCGGCTACAA ATCCGAATCG
1801 GCAGTCGCCA TCGGTACCGG CTTCCGCTTT ACCGAAAACT TTGCCGCCAA
1851 AGCAGGCCTG GCAGTCGGCA CTTGTCGGG TTCTTCCGCA GCCTACCATG
1901 TCGGCGTCAA TTACGAGTGG CTCGAGCACC ACCACCACA CCACTGA

```

```

1 MVAADIGAGL ADALTAFLDH KDKGLQSLTL DQSVKNEKL KLAAGAEKT
51 YNGDSLNTG KLNKDKVSRF DFIRQIEVDG QLITLESSEF QVYKQSHSAL
101 TAFQTEQIQD SEHSGKMWAK RQFRIGDLAG EHTSFDEKLE GGRATYRGTA
151 FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
201 VISGSVLYNQ AEKGSYSLGI FGGKAQEVAG SAEVKTVNGI RHIGLAAKQL
251 EGGGGTGSAT NDDDVKAAT VAIAAAYNNG QEINGFKAGE TIYDIDEDGT
301 ITKRDATAAD VEADDFKGLG LKKVVTNLTK TVNENKQNV DAKVKAASEHI
351 EKLTTKLADT DAALADTDA LDATTNALNK LGENTTTFAR ETKTNIVKID
401 EKLEAVADTV DEHAEAFNDI ADSLDETNTK ADEAVKTANE AKQFAEHTKQ
451 NVDKVKAAE TAAGKAEAAA GTANTAADKA EAVAANKVTDI KADLATNKDN
501 IAKKANSADV YTRHESDSKF VRIDGLNATT EKLDTRLASA EKSIADHDTR
551 LNLGLDKTVSD LRKETRQGLA BQAALSGLFQ FYNVGRFNV TAAVGGYKSES
601 AVAIGTGFRF TENFAAKAGV AVGTSSGSSA AYHVG VNYEW LEHHHHHH*

```

AG741-961c

```

1 ATGGTCGCCG CCGACATCGG TCGGGGGCTT GCCGATGCAC TAACCGCACC
51 GCTCGACCAT AAAGACAAAG GTTTGCAGTC TTTGACGCTG GATCAGTCCG
101 TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAACT
151 TATGGAAACG GTGACAGCCT CAATACGGGC AAATGAAGA ACGACAAGGT
201 CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA
251 CCTTGGAGAG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
301 ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAGAT
351 GGTTCGGA AA CGCCAGTTCA GAATCGGCGA CATAGCGGCG GAACATACAT
401 CTTTGTGACAA GCTTCCCGAA GGCGGCAGGG CGACATATCG CGGGACGGCG
451 TTCGGTTCAG ACGATGCCGG CGGAAACTG ACCTACACCA TAGATTTTCG
501 CGCAAGCAG GGAACCGCA AAATCGAACA TTGAAATCG CCAGAACTCA
551 ATGTCGACCT GGCCGCCGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
601 GTCATCAGCG GTTCCGTCTT TTACAACCAA GCCGAGAAAG GCAGTTACTC
651 CCTCGGTATC TTTGGCGGAA AAGCCCAGGA AGTTGCCGGC AGCGCGGAAG
701 TGA AAACCGT AAACGGCATA CGCCATATCG GCCTTGCCGC CAAGCAACTC
751 GAGGGTGGCG GAGGCACTGG ATCCGCCACA AACGACGACG ATGTTAAAAA

```

```

801 AGCTGCCACT GTGGCCATTG CTGCTGCCTA CAACAATGGC CAAGAAATCA
851 ACCGTTTCAA AGCTGGAGAG ACCATCTACG ACATTGATGA AGACGGCACA
901 ATTACCAAAA AAGACGCAAC TGCAGCCGAT GTTGAAGCCG ACBACTTTAA
951 AGGTCTGGGT CTGAAAAAAG TCGTGACTAA CCTGACCATA ACCGTCAATG
1001 AAAACAAACA AAACGTCGAT GCCAAAGTAA AAGCTGCAGA ATCTGAAATA
1051 GAAAAAGTTAA CAACCAAGTT AGCAGACACT GATGCCGCTT TAGCAGATAC
1101 TGATGCCGCT CTGGATGCAA CCACCAACGC CTTGAATAAA TTGGGAGAAA
1151 ATATAACGAC ATTTGCTGAA GAGACTAAGA CAAATATCGT AAAAAATTGAT
1201 GAAAAATTAG AAGCCGTGCG TGATACCGTC GACAAGCATG CCGAAGCATT
1251 CAACGATATC GCCGATTCTT TGGATGAAAC CAACACTAAG GCAGACGAAG
1301 CCGTCAAAAC CGCCAATGAA GCCAAACAGA CGGCCGAAGA AACCACAAAC
1351 AACGTCGATG CCAAAGTAAA AGCTGCAGAA ACTGCAGCAG GCAAAGCCGA
1401 AGCTGCCGCT GGCACAGCTA ATACTGCAGC CGACAAGGCC GAAGCTGTCG
1451 CTGCAAAAGT TACCGACATC AAAGCTGATA TCGCTACGAA CAAAGATAAT
1501 ATTGCTAAAA AAGCAAACAG TGCCGACGTG TACACCAGAG AAGAGTCTGA
1551 CAGCAATTTT GTCAGAAATG ATGGTCTGAA CGCTACTACC GAAAAATTGG
1601 ACACACGCTT GCGTTCTGCT GAAAAATCCA TTGCCGATCA CGATACTCGC
1651 CTGAACGGTT TGGATAAAAC AGTGTCAAGC CTGCGCAAAG AAACCCGCCA
1701 AGGCTTTGCA GAACAAGCCG CGCTCTCCGG TCTGTTCCAA CCTTACAACG
1751 TGGGTCTCGA GCACCACCAC CACCACCAC GA

```

```

1 MVAADIGAGL ADALTAFLDH KDKGLQSLTL DQSVRKNEKL KLAAGGAERT
51 YGNGDSLNTG KLNKDKVSRF DFIRQIEVDG QLITLESSEF QVYKQSHSL
101 TAFQTEBQID SEHSGKMVAK RQFRIGDIAG KHTSFDELPB GGRATYRGT
151 FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS FELNVDLAAA DIKPDGKRHA
201 VISGSVLYNQ AEKGSYSLGI FGGKAQEVAG SAEVKTVNGI RHIGLAARQL
251 EGGGGTGSAT NDDDVKEAAT VAIAAAYNNG QHNGFKAGE TTYDIDHDGT
301 ITKKDATAAD VEADDFKGLG LKKVVTNLTK TVNENKQNV D AKVKAARSEI
351 EKLTTKLADT DAALADTDAA LDATTNALNK LGENITTFAB STKTNIVKID
401 EKLEAVADTV DKHAEAFNDI ADSLDETNTK ADEAVKTANE AKQTAEETKQ
451 NVDKVKAAE TAAGKAAAAA GTANTAADKA EAVAAKVTDI KADIATNKDN
501 IAKKANSAADV YTRERSOSKF VRIDGLNATT EKLDTRLASA EKSIAHDHTR
551 LNGLDKTVSD LRKETRQGLA EQAALSGLFQ FYNVGLRHHH HHH*

```

AG741-983

```

1 ATGGTCCGCG CCGACATCGG TGCGGGGCTT GCCGATGCAC TAACCGCACC
51 GCTCGACCAT AAAGACAAAG GTTTGCAGTC TTTGAAGCTG GATCAGTCCG
101 TCAGGAAAAA CGAGAAACTG AAGCTGGCGG CACAAGGTGC GGAAAAAACT
151 TATGGAAACG GTGACAGCCT CAATACGGGC AAATTGAAGA ACGACAAGGT
201 CAGCCGTTTC GACTTTATCC GCCAAATCGA AGTGGACGGG CAGCTCATTA
251 CCTTGAGAGG TGGAGAGTTC CAAGTATACA AACAAAGCCA TTCCGCCTTA
301 ACCGCCTTTC AGACCGAGCA AATACAAGAT TCGGAGCATT CCGGGAAGAT
351 GGTTCGGAAG CGCCAGTTCA GAATCGGCGA CATAGCGGCG GAACATACAT
401 CTTTTGACAA GCTTCCCGAA GCGCGCAGGG CGACATATCG CCGGACGGCG
451 TTCGGTTCAG ACGATGCCGG CGGAAACTG ACCTACACCA TAGATTTCCG
501 CGCCAGCAGG GGAACCGCA AAATCGAACA TTTGAAATCG CCAGAACTCA
551 ATGTCGACCT GCGCGCGGCC GATATCAAGC CGGATGGAAA ACGCCATGCC
601 GTCATCAGCG GTTCCGTCCT TTACAACCAA GCCGAGAAG GCAGTTACTC
651 CCTCGGTATC TTTGGCGGAA AAGCCCAGGA AGTTGCCGCG AGCGCGGAAG
701 TGAAAACCGT AAACGGCATA CGCCATATCG GCCTTGCCGC CAAGCAACTC
751 GAGGGATCCG GCGGAGGCGG CACTTCTGCG CCCGACTTCA ATGCAGGCGG
801 TACCGGTATC GGCAGCAACA GCAGAGCAAC AACAGCGAAA TCAGCAGCAG
851 TATCTTACGC CGGTATCAAG AACGAAATGT GCAAAGACAG AAGCATGCTC
901 TGTGCCGGTC GGGATGACGT TGCGGTTACA GACAGGGATG CCAAAATCAA
951 TGCCCCCCCC CCGAATCTGC ATACCGGAGA CTTTCCAAAC CCAATGACG
1001 CATAACAAGAA TTTGATCAAC CTCAAACCTG CAATTGAAGC AGGCTATACA
1051 GGACGCGGGG TAGAGGTAGG TATCGTTCGAC ACAGGCGAAT CCGTCGGCAG
1101 CATATCCTTT CCCGAACTGT ATGGCAGAAA AGAACACGGC TATAACGAAA
1151 ATTACAAAAA CTATACGGCG TATATGCGGA AGGAAGCGCC TGAAGACGGA
1201 GCGGTAAAG ACATTGAAGC TTCTTTTCGAC GATGAGGCCG TTATAGAGAC
1251 TGAAGCAAAG CCGACGGATA TCCGCCACGT AAAAGAAATC GGACACATCG
1301 ATTTGGTCTC CCATATTATT GCGGGCGGTT CCGTGGACGG CAGACCTGCA
1351 GCGGTATTG CCGCGATGC GACGCTACAC ATAATGAATA CGAATGATGA
1401 AACCAAGAAC GAAATGATGG TTGCAGCCAT CCGCAATGCA TGGGTCAAGC
1451 TGGGCGAAGC TGGCGTGGCG ATCGTCAATA ACAGTTTGG AACACATCG
1501 AGGCGAGGCA CTGCCGACCT TTTCCAAATA GCCAATTCGG AGGAGCAGTA

```

```

1551 CCGCCAAGCG TTGCTCGACT ATTCCGGCGG TGATAAAACA GACGAGGGTA
1601 TCCGCCTGAT GCAACAGAGC GATTACGGCA ACCTGTCCTA CCACATCCGT
1651 AATAAAAACA TGCTTTTCAT CTTTTCGACA GGCAATGACG CACAAGCTCA
1701 GCCCAACACA TATGCCCTAT TGCCATTTTA TGAAAAAGAC GCTCAAAAAG
1751 GCATTATCAC AGTCGCAGGC GTAGACCGCA GTGGAGAAAA GTTCAAACGG
1801 GAAATGTATG GAGAACCGGG TACAGAACCG CTTGAGTATG GCTCCAACCA
1851 TTGCGGAATT ACTGCCATGT GGTGCCTGTC GGCACCTAT GAAGCAAGCG
1901 TCCGTTTCAC CCGTACAAAC CCGATTCAAA TTGCCGGAAC ATCCTTTTCC
1951 GCACCCATCG TAACCGGCAC GGCGCCTCTG CTGCTGCAGA AATACCCGTG
2001 GATGAGCAAC GACAACCTGC GTACCACGTT GCTGACGACG GCTCAGGACA
2051 TCGGTGCACT CGGCGTGGAC AGCAAGTTCG GCTGGGGACT GCTGGATGCG
2101 GGTAAAGCCA TGAACGGACC CGCGTCCTTT CCGTTCGGCG ACTTTACCGC
2151 CGATACGAAA GGTACATCCG ATATTGCCTA CTCCTTCCGT AACGACATTT
2201 CAGGCACGGG CGGCCTGATC AAAAAAGGCG GCAGCCAACT GCAACTGCAC
2251 GGCAACAACA CCTATACGGG CAAAACCATT ATCGAAGGCG GTTCGCTGGT
2301 GTTGATACGG AACAACAAAT CGGATATGCG CGTCGAAACC AAAGGTGCGC
2351 TGATTTATAA CGGGGCGGCA TCCGCGGGCA GCCTGAACAG CGACGGCATT
2401 GTCTATCTGG CAGATACCGA CCAATCCGGC GCAAACGAAA CCGTACACAT
2451 CAAAGGCAGT CTGCAGCTGG ACGGCAAAGG TACGCTGTAC ACACGTTTGG
2501 GCAAACCTGCT GAAAGTGGAC GGTACGGCGA TTATCGGCGG CAAGCTGTAC
2551 ATGTCGGCAC CGGCAAGGG GGCAGGCTAT CTCAACAGTA CCGACGACG
2601 TGTTCCTCTC CTGAGTGCCG CAAAATCGG GCAGGATTAT TCTTCTTCA
2651 CAAACATCGA AACCGACGGC GGCCTGCTGG CTTCCCTCGA CAGCGTCGAA
2701 AAAACAGCGG GCAAGTGAAG CGACACGCTG TCCTATATAG TCCGTGCGCG
2751 CAATCGGGCA CGGACTGCTT CGGCAGCGGC ACATTCCGCG CCCGCGGCTC
2801 TGAACACGCG CGTAGAACAG GCGGCGAGCA ATCTGAAAAA CCTGATGGTC
2851 GAACCTGGATG CTTCCGAATC ATCCGCAACA CCGAGACGG TTGAAACTGC
2901 GGCAGCCGAC CGCACAGATA TGCCGGGCAT CCGCCCCTAC GGCACAACTT
2951 TCCGCGCAGC GGCAGCCGTA CAGCATGCGA ATGCCGCCGA CGGTGTACGC
3001 ATCTTCAACA GTCTCGCGC TACCGTCTAT GCCGACAGTA CCGCGGCCCA
3051 TGCCGATATG CAGGGACGCC GCCTGAAAGC CGTATCGGAC GGGTTGGACC
3101 ACAACGGCAC GGTCTGCGC GTCATCGCGC AAACCCAACA GGACGGTGA
3151 ACGTGGGAAC AGGGCGGTGT TGAAGGCAAA ATGCGCGGCA GTACCCAAAC
3201 CGTCCGCATT GCCGCGAAAA CCGGCGAAAA TACGACAGCA GCCCGCACAC
3251 TGGGCATGGG ACGCAGCACA TGGAGCGAAA ACAGTGCAA TGCAAAAACC
3301 GACAGCATTG GTCTGTTTGC AGGCATACGG CACGATGCGG GCGATATCGG
3351 CTATCTCAAA GGCCTGTTCT CCTACGGACG CTACAAAAAC AGCATCAGCC
3401 GCAGCACCGG TCGGACGAA CATGCGGAAG GCAGCGTCAA CCGCACGCTG
3451 ATGCAGCTGG GCGCACTGGG CGGTGTCAAC GTTCCGTTTG CCGCAACGGG
3501 AGATTTGACG GTCGAAGGCG GTCTGCGCTA CGACCTGCTC AAACAGGATG
3551 CATTCCGCGA AAAAGGCAGT GCTTTGGGCT GGAGCGGCAA CAGCCTCACT
3601 GAAGGCACGC TGGTCGACT TGGTGGGCTG AAGCTGTGCG AACCTTGAG
3651 CGATAAAGCC GTCCTGTTTG CAACGGCGGG CGTGGAAACG GACCTGAACG
3701 GACGCGACTA CACGTAACG GCGGCGTTTA CCGGCGCGAC TGCAGCAACC
3751 GGCAAGACGG GGCACGCAA TATGCCGCAC ACCCGTCTGG TTGCCGCGCT
3801 GGGCGCGGAT GTCGAATTCG GCAACGGCTG GAACGGCTTG GCACGTTACA
3851 GCTACGCCGG TTCCAAACAG TACGGCAACC ACAGCGGACG AGTCGGCGTA
3901 GGCTACCGGT TCCTCGAGCA CCACCACCAC CACCACTGA

```

```

1 MVAADIGAGL ADALTAFLDH KDKGLQSLTL DQSVRENEKL KLAAQGAET
51 YNGDSLNTG KLNKDKVSRF DFIRQIEVDG QLITLESGEF QVYKQSHSAL
101 TAFQTEBQID SEHSGKMAK RQFRIGDIAG EHTSFDKLPZ GGRATYRGA
151 FGSDDAGGKL TYTIDFAAKQ GNGKIEHLKS PELNVDLAAA DIKPDGKRHA
201 VISGSVLVYQ AEKGSYSLGI FGGKAQEVAG SAEVKTVNGI RHIGLAAKQL
251 ESGGGGTS PDFNAGGTGI GSNRATTAK SAAVSYAGIK NEMCKDRSML
301 CAGRDDVAVT DRDAKINAPP PNLHTGDFPN PNDAYKNLIN LKPAIEAGYT
351 GRGVEVGIVD TGESVGSISF PELYGRKEHG YNENYKNYTA YMRKEAPEDG
401 GGDIEASFD DEAVIEFEAK PTDIRHVKEI GHIDLVSII GGRSVDGRPA
451 GGIAPDATLH IMNTNDETKN EMMVAAIRNA WVKLGERRVR IVNNSFGTTS
501 RAGTADLFQI ANSEBQYRQA LLDYSGGDKT DEGIRLMQQS DYGNLSYHIR
551 NKMMLFIFST GNDQAQPFNT YALLPFYEKD AQKGIIIVAG VDRSGEKPKR
601 EMYGEPGTEP LEYGSNHCIG TAMNCLSAFY EASVRFTRTN PIQIAGTSFS
651 APIVTGTAAL LLQKYPMSN DMLRTTLLTT AQDIGAVGVD SKFGWGLLDA
701 GKAMNGPASF PFGDFTADTK GTSDIAYSFR NDISGTGLI KKGSQQLQLH
751 GNNYTYTGKI IEGGSLVLYG NNXSDMRVET KQALTYNGAA SGGSLNSDGI
801 VYLADTDQSG ANETVHIKGS LQLDGKGTLY TRLGKLLAVD GTAIIGKLY
851 MSARGKGAGY LNSTGRVVPF LSAKIGQDY SFFTNIETDG GLLASLDSVE

```

```

901   KTAGEGDTL  SYTVRRGNAA  RTASAAAHSA  PAGLKHAVEQ  GGSNLENLMV
951   ELDAESSAT  PETVETAAAD  RTDMPGIRPY  GATFRAAAAV  QHANAADGVR
1001  IFNSLAATVY  ADSTAAHADM  QGRRLKAVSD  GLDHNGTGLR  VIAQTQQDGG
1051  TWEQGGVEGK  MRGSTQTVGI  AAKTGENTTA  AATLGMGRST  WSENSANAKT
1101  DSISLPAFIR  HDAGDIGYLK  GLFSYGRYKN  SISRSTGADE  HAEGSVNGTL
1151  MQLGALGGVN  VFPAATGDLT  VEGGLRYDLL  KQDAFAEKGS  ALWSGNSLT
1201  EGTLVGLAGL  KLSQPLSDKA  VLFATAGVER  DLNGRDYTVT  GGFTGATAAT
1251  GKTGARMPH  TRLVAGLGAD  VEFNGWNGL  ARYSYAGSKQ  YGNHSGRVGV
1301  GYRFLHHHH  HH*

```

AG741-ORF46.1

```

1   ATGGTCGCCG  CCGACATCGG  TCGGGGGCTT  GCCGATGCAC  TAACCGCACC
51  GCTCGACCAT  AAAGACAAAG  GTTTGCAGTC  TTGACGCTG  GATCAGTCCG
101  TCAGGAAAAA  CGAGAACTG  AAGCTGGCGG  CACAAGGTGC  GGAAAAAACT
151  TATGGAACG  GTGACAGCCT  CAATACGGGC  AAATTGAAGA  ACGACAAGGT
201  CAGCCGTTTC  GACTTTATCC  GCCAAATCGA  AGTGGACGGG  CAGCTCATTA
251  CCTTGGAGAG  TGGAGAGTTC  CAAGTATACA  AACAAAGCCA  TTCCGCCTTA
301  ACCCGCTTTC  AGACCGAGCA  AATACAAGAT  TCGAGCATT  CCGGGAAGAT
351  GGTTCGAAA  CGCCAGTTCA  GAATCGCGCA  CATAGCGGGC  GAACATACAT
401  CTTTGTGACA  GCTTCCCGAA  GCGGCGAGGG  CGACATATCG  CGGGACGGCG
451  TTCGGTTCAG  ACGATGCCGG  CGGAAACTG  ACCTACACCA  TAGATTTCGC
501  CGCCAAGCAG  GGAAACGGCA  AAATCGAACA  TTTGAAATCG  CCAGAACTCA
551  ATGTCGACCT  GCGCGCCGCC  GATATCAAGC  CGGATGGAAA  ACGCCATGCC
601  GTTCATCAGC  GTTCGGTCC  TTACACCAA  GCGGAGAAAG  GCAGTTACTC
651  CCTCGGTATC  TTTGGCGGAA  AAGCCCAGGA  AGTTGCCGGC  AGCGCGGAAG
701  TGAAAAACCGT  AAACGGCATA  CGCCATATCG  GCCTTGCCGC  CAAGCAACTC
751  GACGGTGGCG  GAGGCACTGG  ATCCTCAGAT  TTGGCAAACG  ATTCTTTTAT
801  CCGGCAGGTT  CTCGACCGTC  AGCATTTGGA  ACCCGACGGG  AAATACCAAC
851  TATTCCGCAG  CAGGGGGGAA  CTTGCCGAGC  GCAGCGGCCA  TATCGGATTG
901  GGAAAAATAC  AAAGCCATCA  GTTGGGCAAC  CTGATGATTC  AACAGGCGGC
951  CATTAAAGGA  AATATCGGCT  ACATTGTCCG  CTTTTCGGAT  CACGGGCACG
1001  AAGTCCATTC  CCCCTTCGAC  AACCATGCCT  CACATTCCGA  TTCTGATGAA
1051  GCGGCTAGTC  CCGTTGACGG  ATTTAGCCTT  TACCGCATCC  ATTGGGACGG
1101  ATACGAACAC  CATCCCGCCG  ACGGCTATGA  CGGGCCACAG  GCGGCGGCT
1151  ATCCCGCTCC  CAAAGGCGCG  AGGGATATAT  ACAGCTACGA  CATAAAGGC
1201  GTTGCCTCAA  ATATCCGCCT  CAACCTGACC  GACAACCGCA  GCACCGGACA
1251  ACGGCTTGCC  GACCGTTTCC  ACAATGCCGG  TAGTATGCTG  ACGCAAGGAG
1301  TAGGCGACGG  ATTCAAACGC  GCCACCGGAT  ACAGCCCCGA  GCTGGACAGA
1351  TCGGGCAATG  CCGCCGAAGC  CTTCAACGGC  ACTGCAGATA  TCGTTAAAAA
1401  CATCATCGGC  GCGGCAGGAG  AAATTGTGG  CGCAGGCGAT  GCCGTGCAGG
1451  GCATAAGCGA  AGGCTCAAAC  ATTGCTGTCA  TGCACGGCTT  GGGTCTGCTT
1501  TCCACCGAAA  ACAAGATGGC  GCGCATCAAC  GATTGGCAG  ATATGCGGCA
1551  ACTCAAAGAC  TATGCCGCG  CAGCCATCCG  CGATTGGGCA  GTCCAAAACC
1601  CCAATGCCGC  ACAAGGCATA  GAAGCCGTCA  GCAATATCTT  TATGGCAGCC
1651  ATCCCCATCA  AAGGATTGG  AGCTGTTCCG  GGAAATACG  GCTTGGGCGG
1701  CATCACGGCA  CATCTATCA  AGCGGTCCGA  GATGGGCGCG  ATCGCATGTC
1751  CGAAAGGGAA  ATCCGCGGTC  AGCGACAATT  TTGCCGATGC  GGCATACGCC
1801  AAATACCGT  CCCCTTACCA  TTCCCGAAAT  ATCCGTTCAA  ACTTGGAGCA
1851  GCGTTACGGC  AAAGAAAACA  TCACCTCCTC  AACCCTGCCG  CCGTCAAACG
1901  GCAAAAATGT  CAAACTGGCA  GACCAACGCC  ACCCGAAGAC  AGGCGTACCG
1951  TTTGACGGTA  AAGGTTTCC  GAATTTTGAG  AAGCAGGTGA  AATATGATAC
2001  GCTCGAGCAC  CACCACCACC  ACCACTGA

```

```

1   MVAADIGAGL  ADALTAFLDH  KDKGLQSLTL  DQSVRKNEKL  KLAAQGAZKT
51  YNGDSLNTG  KLNKDVSRF  DFIRQIEVDG  QLITLESGEF  QVYKQSHSAL
101  TAFQTEQIQD  SEHSGKMVA  RQFRIGDIAG  EHTSFDELP  GGRATYRGT
151  FGSDDAGGKL  TYTIDFAAKQ  GNGKIEHLKS  PELNVDLAA  DIKPDGKRHA
201  VISGSLVLYN  AEKGSYSLGI  FGGKAQEVAG  SAEVKTVNGI  RHIGLAAKQL
251  DGGGTGSSD  LANDSFIRQV  LDRQHFEFDG  KYHLFGSRGE  LAERSGHIGL
301  GKIQSHQLGN  LMIQQAALIK  NIGYIVRFS  HGHEVHSFFD  NEASHSDSDE
351  AGSPVDGFS  YRIHWDGYEH  HPADGYDGPQ  GGGYPAPKGA  RDIYSYDIK
401  VAQNIRLNLT  DNRSTGQRLA  DRPHNAGSML  TQGVGDGFKR  ATRYSPFLDR
451  SGNAABAFNG  TADIVKNIIG  AAGEIVGAGD  AVQGISSEGS  LAVMHGLGLL
501  STENMARIN  DLADMAQLKD  YAAAAIRDWA  VQNPNAAGGI  EAVSNIFMAA
551  IPIKGIGAVR  GKYGLGGITA  HPIKRSQMG  IALPKGKSAV  SDNFADAAYA
601  KYPSFYHSRN  IRSNLEQRYG  KENITSSTVP  PSNGKNVKLA  DQREPKTGVP
651  FDGKGFPNFE  KHVKYDTLKH  HHHHH*

```

实施例 16-与 287/△G287 的 C-末端融合(‘杂交体’)

按照本发明，两种蛋白质 A 和 B 的杂交体可能是 $\text{NH}_2\text{-A-B-COOH}$ 或 $\text{NH}_2\text{-B-A-COOH}$ 。用在 919、953 和 ORF46.1 的 C-端(以 '287-His'形式)或 N-末端(以 ΔG287 形式-如上所示的序列)的蛋白质 287 研究了这种差异的影响。使用一组菌株，包括同源菌株 2996。用 FCA 作为佐剂：

	287 & 919		287 & 953		287 & ORF46.1	
菌株	$\Delta\text{G287-919}$	919-287	$\Delta\text{G287-953}$	953-287	$\Delta\text{G287-46.1}$	46.1-287
2996	128000	16000	65536	8192	16384	8192
BZ232	256	128	128	<4	<4	<4
1000	2048	<4	<4	<4	<4	<4
MC58	8192	1024	16384	1024	512	128
NGH38	32000	2048	>2048	4096	16384	4096
394/98	4096	32	256	128	128	16
MenA (F6124)	32000	2048	>2048	32	8192	1024
MenC (BZ133)	64000	>8192	>8192	<16	8192	2048

5

通常用 N-末端的 297(以 ΔG 的形式)观察到较好的杀菌滴定度。

当融合于蛋白质 961[$\text{NH}_2\text{-}\Delta\text{G287-961-COOH}$ -上述可示序列]时，得到的蛋白质是不溶的且必须变性和复性以用于纯化。复性后，发现 50%蛋白质仍不溶解。比较可溶性和不溶性蛋白质，用可溶性蛋白质获得的杀菌滴定度好得多

10 (FCA 作为佐剂)：

	2996	BZ232	MC58	NGH38	F6124	BZ133
可溶性	65536	128	4096	>2048	>2048	4096
难溶性	8192	<4	<4	16	n.d.	n.d.

然而，用明矾佐剂替代可以改善难溶形式蛋白质的滴定度：

难溶性	32768	128	4096	>2048	>2048	2048
-----	-------	-----	------	-------	-------	------

实施例 17-N-末端融合('杂交体')于 287

15 将蛋白质 287 以全长、含 C-末端 His-标记、或无其前导肽但而包含 C-末端 His-标记表达的水平相当低。用 N-末端 GST-融合实现较好的表达。

或者将 GST 用作 N-末端融合配体，将 287 置于蛋白质 919 的 C-末端 ('919-287')、蛋白质 953 的 C-末端 ('953-287')和蛋白质 ORF46.1 的 C-末端 ('ORF46.1-287')。在这两种情况中，除去前导肽，且杂交体是直接的符合读框

的融合。

为了产生 953-287 杂交体，设计各序列的前导肽的正向引物下游以除去这两种蛋白质的前导肽；除去 953 反向引物中的终止密码子序列，但包含在 287 反向引物中。对 953 基因而言，用于扩增的 5'和 3'引物分别包括 NdeI 和 BamHI 5 限制酶切位点，而对 287 基因的扩增而言，5'和 3'引物分别包括 BamHI 和 XhoI 限制酶切位点。在这种方法中，使用 NdeI-BamHI(用于克隆第一个基因)和随后的 BamHI-XhoI(用于克隆第二基因)可以实现 pET21b+中两个基因的顺序定向克隆。

通过将编码 287 的成熟部分的序列克隆入 pET21b+中的 919-His 克隆的 3' 10 端的 XhoI 位点中，获得 919-287 杂交体。设计用于扩增 287 基因的引物，从而在 PCR 片段的 5'端引入 SalI 限制酶切位点，并在 3'端引入 XhoI 位点。由于由 SalI 和 XhoI 限制酶产生的粘性末端是相容的，可以将用 SalI-XhoI 消化的 287PCR 扩增产物插入由 XhoI 切割的 pET21b-919 克隆中。

类似地获得 ORF46.1-287 杂交体。

15 将针对杂交蛋白产生的抗体的杀菌效力（同源菌株）与针对组分抗原的简单混合物产生的抗体的杀菌效力相比较：

	与 287 的混合物	与 287 的杂交体
919	32000	16000
953	8192	8192
ORF46.1	128	8192

还获得了 919-287 和 953-287 针对异源 MenB 菌株和血清型 A 和 C 的杀菌活性数据：

菌株	919		953		ORF46.1	
	混合物	杂交体	混合物	杂交体	混合物	杂交体
MC58	512	1024	512	1024	-	1024
NGH38	1024	2048	2048	4096	-	4096
BZ232	512	128	1024	16	-	-
MenA (F6124)	512	2048	2048	32	-	1024
MenC (C11)	>2048	n.d.	>2048	n.d.	-	n.d.
MenC (BZ133)	>4096	>8192	>4096	<16	-	2048

还构建了 ORF46.1 和 919 的杂交体。最佳的结果(高四倍的滴定度)是用于 N-末端的 919 实现的。

还测试了杂交体 919-519His、ORF97-225His 和 225-ORF97His。它们得到中等 ELISA 滴定度和杀菌的抗体应答。

5

实施例 18-ORF4 的前导肽

如上所示,可以将 ORF4 的前导肽融合于其它蛋白质(如蛋白质 287 和 919)的成熟序列。其可以指导大肠杆菌中的脂质化。

10 实施例 19-564 中的结构域

蛋白质'564'非常大(2073aa), 且其难以克隆并以完整形式表达。为了便于表达, 如图 8 所示(按 MC58 序列)将该蛋白质分为四个结构域:

结构域	A	B	C	D
氨基酸	79-360	361-731	732-2044	2045-2073

这些结构域显示出如下同源性:

• 结构域 A 显示与其它细菌毒素同源:

- 15 gb | AAG03431.1 | AE004443-9 可能血凝素[绿脓假单胞菌](38%)
 gb | AAC31981.1 | (139897) HecA[菊果胶杆菌](45%)
 emb | CAA36409.1 | (X52156) 丝状血凝素[百日咳博德特氏菌](31%)
 gb | AAC79757.1 | (AF057695) 大量上清液蛋白质 1[杜氏嗜血菌](26%)
 gb | AAA25657.1 | (M30186) HpmA 前体[奇异变形菌](29%)

20 • 结构域 B 显示为非同源的, 且是对 564 特异性的。

• 结构域 C 显示为与以下同源:

- gb | AAF84995.1 | AE004032 HA-样分泌的蛋白质[苛养木杆菌](33%)
 gb | AAG05850.1 | AE004673 假说蛋白质[铜绿假单胞菌](27%)
 gb | AAF68414.1AF237928 推定的 FHA[多杀巴斯德氏菌](23%)
 25 gb | AAC79757.1 | (AF057695) 大量上清液蛋白质 1[杜氏嗜血菌](23%)
 pir | | S21010 FHA B 前体[百日咳博德特氏菌](20%)

• 结构域 D 显示与前体细菌毒素同源:

gb | AAF84995.1 | AE004032_14 HA-样分泌的蛋白质[苛养木杆菌](29%)

用 MC58 菌株序列, 以 GST-融合(未纯化)和 his-标记的蛋白质形式获得 564ab 良好的细胞内表达; 该结构域-对还以脂蛋白形式表达, 其表现出在外膜/上清液组分中有中等表达。

- 5 当以 his-标记的产物(未纯化)表达时, b 结构域显示中等细胞内表达, 而以 GST-融合则表达良好。

以 GST-融合表达时 c, 结构域显示良好的细胞内表达, 但是不溶的。以 his 标记的产物(未纯化)表达时, d 结构域显示中等细胞内表达。以 GST-融合体表达时, cd 蛋白质结构域对表现中等细胞内表达(未纯化)。

- 10 用 c 结构域和 bc 对观察到良好的杀菌试验滴定度。

实施例 20-919 前导肽

在以上的实施例 1 中讨论了 919 的 20mer 前导肽:

MKKYLFRAL YGIAAAILAA

- 15 如实施例 1 所示, 该前导肽的缺失改进了异源表达(如用 ORF4 前导肽替代该前导肽一样)。通过将编码序列融合于摩氏摩根氏菌(*Morganella morganii*)的 PhoC 报道基因研究了 919 前导肽对表达的影响[Thaller 等, (1994)Microbiology 140: 1341-1350]。将该构建物克隆入 NdeI 和 XhoI 位点间的 pET21-b 质粒中(图 9):

20

1	MKKYLFRAL YGIAAAILAA	AIPAGNDAT KPDLYYLKNE QAIDSLKLLP
51	PPFEVGSIQF LNDQAMYKKG	RMLENTERGK QAQADADLAA GGVATAFSGA
101	FGYPITERKDS PELYKLLTMM	IEDAGOLATR SAKHYMRIR PFAPYGTETC
151	NTKDQKLLST NGSYPSGHTS	IGNATALVLA EVNPAHQDAI LERGTYLQGS
201	RVICGYHMQS DVDAARIVGS	AAVATLHSDR AFQAQLAKAK QEFAQKSR*

该质粒的 PhoC 的表达水平比相同的但包含天然 PhoC 信号肽的构建物中发现的低 >200 倍。即使用大肠杆菌 Plac 启动子替代 T7 启动子后也得到相同的结果。这表明 919 前导序列对表达的影响并非取决于所用的启动子。

- 25 为了研究这些结果是否是由于 919 信号肽核苷酸序列(二级结构形式, 对 RNA 酶等敏感)的某些特性或存在该信号肽引起的蛋白质不稳定性导致的, 产生了许多突变体。所用的方法是通过克隆含简并密码子的合成接头替代 919 信号肽序列的核苷酸。用这种方法获得含核苷酸和/或氨基酸取代的变体。

使用两种不同的接头，设计 919 信号肽序列的两个不同区域中产生突变，即在第一 19 碱基对(L1)中和碱基 20-36 之间(S1)。

L1: 5' T ATG AAA/g TAc/t c/tTN TTt/c a/cGC GCC GCC CTG TAC GGC ATC GCC GCC
GCC ATC CTC GCC GCC GCG ATC CC 3'
S1: 5' T ATG AAA AAA TAC CTA TTC CGa/g GCN GCN c/tTa/g TAc/t GGc/g ATC GCC
GCC GCC ATC CTC GCC GCG GCG ATC CC 3'

获得的一些突变体的序列排序如下。

5 L1 变体:

```
9L1-a ATGAAGAAGTACCTTTTCAGCGCGGCC-----
9L1-e ATGAAAAAATACCTTTTCGCGCGCGCC-----
9L1-d ATGAAAAAATACCTTTTCGCGCGCGCC-----
9L1-f ATGAAAAAATATCTCTTTAGCGCGCGCCTGTACGGCATCGCCGCCCATCTCTCGCGGCC
919sp ATGAAAAAATACCTATTCGCGCGCGCCCTGTACGGCATCGCCGCCCATCTCTCGCGGCC

9L1a MKKYLFSAA-----
9L1e MKKYFFRAA-----
9L1d MKKYFFRAA-----
9L1f MKKYLFSAALYGIAAAILAA
919sp MKKYLFRALYGIAAAILAA (即，天然信号肽)
```

S1 突变体:

```
9S1-e ATGAAAAAATACCTATTC.....ATCGCGCGCGCCATCTCTCGCGGCC
9S1-c ATGAAAAAATACCTATTCGAGCTGCCCAATACGGCATCGCCGCCCATCTCTCGCGGCC
9S1-b ATGAAAAAATACCTATTCGGGGCGGCCAATACGGCATCGCCGCCCATCTCTCGCGGCC
9S1-i ATGAAAAAATACCTATTCGGGGCGGCTTTGTACGGCATCGCCGCCCATCTCTCGCGGCC
919sp ATGAAAAAATACCTATTCGCGCGCGCCCTGTACGGCATCGCCGCCCATCTCTCGCGGCC

9S1e MKKYL.....IAAAILAA
9S1c MKKYLFRAAQYGIAAAILAA
9S1b MKKYLFRAAQYGIAAAILAA
9S1i MKKYLFRALYGIAAAILAA
919sp MKKYLFRALYGIAAAILAA
```

如序列排序中所示，分析的的大部分突变体包含由宿主细胞出人意料地产生的符合读框的缺失。

通过用从 L1 和 S1 突变的克隆制备的 DNA 转化大肠杆菌 BL21(DE3)细胞进行突变体的筛选。通过在包含 100 μ g/ml 氨苄青霉素、50 μ g/ml 甲基绿、1mg/ml PDP(酚酞二磷酸)的 LB 平板上划线，筛选高 PhoC 活性的单转化体。在该培养基上产生 PhoC 的细胞变成绿色(图 10)。

15 用 pNPP 作为 PhoC 活性的底物，在液体培养基中对这些突变体产生的 PhoC 进行的定量分析。测定在液体培养基中生长 0、30、90、180 分钟的实变体的细胞提取物和上清液的比活性为：

细胞提取物

	0	30	90	180
c 对照	0,00	0,00	0,00	0,00
9phoC	1,11	1,11	3,33	4,44
9S1e	102,12	111,00	149,85	172,05
9L1a	206,46	111,00	94,35	83,25
9L1d	5,11	4,77	4,00	3,11
9L1f	27,75	94,35	82,14	36,63
9S1b	156,51	111,00	72,15	28,86
9S1c	72,15	33,30	21,09	14,43
9S1l	156,51	83,25	55,50	26,64
phoCwt	194,25	180,93	149,85	142,08

上清液

	0	30	90	180
对照	0,00	0,00	0,00	0,00
9phoC	0,33	0,00	0,00	0,00
9S1e	0,11	0,22	0,44	0,89
9L1a	4,88	5,99	5,99	7,22
9L1d	0,11	0,11	0,11	0,11
9L1f	0,11	0,22	0,11	0,11
9S1b	1,44	1,44	1,44	1,67
9S1c	0,44	0,78	0,56	0,67
9S1l	0,22	0,44	0,22	0,78
phoCwt	34,41	43,29	87,69	177,60

一些突变体产生大量 PhoC，尤其是突变体 9L1a 可以在培养基中分泌 5 PhoC。这是值得关注的，因为该变体的信号肽序列的长度仅为 9 个氨基酸。这是至今最短的信号肽。

实施例 21-Maf-相关蛋白质的 C-端缺失

MafB-相关蛋白质包括 703、ORF46 和 ORF29。

10 MC58 的 730 蛋白质具有如下序列：

```

1  VKPLRLI/TML LAACAVAAAA LIOPALAADL AQDPFITDNA QRQHYEPGGK
51  YHLFGDPRGS VSDRTGKINV IQDYTHQMGV LLIQQANING TIGYETRFSG
101 HGHEEHAPFD NHAADSASKE KGNVDEGFTV YELNWEQHEH HPADAYDGPK
151 GGNYPKPTGA RDEYTYEVNG TARSIKLNPT DTRSIQRIS DNTSNLGSNF
201 SDRADKANRK MFEHNAKLDK WGNSEMFING VAAGALNPFI SAGEALGIGD
251 ILYGTRYAID KAAMRNIAPI PAEGKFAVIG GLGSVAGFEK NTRIAVDKMI
301 QENPNAETV EAVFNVAATA KVAKLAKAAK PGKAAVSGDF ADSYKKKLAL

351 SDSARQLYQN AKYREALDIH YEDLIRKRTD GSSKFINGRE IDAVTNDALI
401 QAKRTISAID KPNFLNQKN RKQIKATIRA ANQQKRAEF WFKYGVHSQV
451 KSYIESKGGI VKTGLGD*

```

前导肽是加下划线的。

730 显示与 ORF46 类似的特性(见以上实施例 8)：

-与 Orf46 一样, 730 序列在 MenB、MenA 和淋球菌之间的保守性仅对 N-末端部分高(>80%)。C-末端(自~340)是高度趋异的。

-其预计的二级结构包含跨越分子中心区域(aa.227-247)的疏水片段。

-大肠杆菌中全长基因的表达产生产量很低的蛋白质。从标记的或未标记的
5 构建物(除去信号肽序列)的表达对宿主细胞有毒性影响。换言之, 细胞质中全长成熟蛋白质的存在对宿主细胞是非常毒的, 而当其转位至周质(由信号肽介导)对细胞生存力没有可检测的影响。730 的这种“细胞内毒性”特别高, 因为只可用 *recA* 遗传背景(大肠杆菌菌株: 用于克隆的 HB101; 用于表达的 HMS174(DE3))以很低的频率得到用于表达无前导肽的 730 的克隆。

10 为了克服这种毒性, 将实施例 8 中用于 ORF46 的类似方法用于 730。得到四种 C-末端截短的形式, 每种都被很好地表达。它们都是从 His-标记的无前导肽 730 的细胞内表达得到的。

形式 A 由成熟蛋白质的 N-末端疏水区域(aa.28-226)构成。将其纯化性作为可溶性 His 标记的产物, 其具有比预计更高的分子量(MW)。

15 形式 B 延伸至血清群之间保守的区域(aa.28-340)的末端。将其纯化为不溶性 His 标记的产物。

在筛选能在菌株 HMS174(DE3)中高水平表达 730-His 克隆的克隆后, 获得命名为 C1 和 C2 的 C-末端截短的形式。简言之, 用含有编码全长成熟 730 蛋白质的 His 标记的序列的 pET21b 质粒转化 *recA* 菌株 HMS174(DE3)。低频率得
20 到转化体, 其显示出两种表型: 大集落和非常小的集落。分析了若干大集落和小集落对 730-His 克隆的表达。只有大集落的细胞超量表达抗 730A 抗体识别的蛋白质。但是, 不同集落中超量表达的蛋白质的分子量各不相同。对其中的两个克隆进行测序表明: 这两种情况中, 在编码 730 的 C 末端区域的序列中发生了大肠杆菌 IS 序列的整合。这两个整合在 C1 中产生了与另一个密码子的符合
25 读框的融合, 在 C2 中产生了与另外 12 个密码子的符合读框的融合(图 11)。产生的 730 “突变体”形式具有如下的序列:

730-C1(IS1 插入引起的-图 11A)

```

1  MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
51  QMGNLLIQQA NINGTIGYET RPSGHGHEEH APFDNHAADS ASEKGNVDE
101 GFTVYRLNWE GHEHPADAY DGPKGNYPK PTGARDEYTY HVNOTARSIX
151 LNPTDTRSIR QRISDNYSNL GSNFSDRADE ANRIMFEHNA KLDKNGWSME
201 FINGVAAGAL NPFISAGEAL GIGDILYGTR YAIIDKAAMRN IAPLPABGKF
251 AVIGGLGSAV GFKNWTRAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
301 KAAKPGKAAV SGDFADSYKK KLALSDSARQ LYQNAKYREA LDHYEDLIR
351 RRTDGSSKFI NGREIDAVTN DALIQAR*

```

由插入产生的其它氨基酸下加下划线。

730-C2(IS5 插入引起的-图 11B)

```

1  MADLAQDPFI TDNAQRQHYE PGGKYHLFGD PRGSVSDRTG KINVIQDYTH
51  QMGNLLIQQA NINGTIGYET RPSGHGHEEH APFDNHAADS ASEKGNVDE
101 GFTVYRLNWE GHEHPADAY DGPKGNYPK PTGARDEYTY HVNOTARSIX
151 LNPTDTRSIR QRISDNYSNL GSNFSDRADE ANRIMFEHNA KLDKNGWSME
201 FINGVAAGAL NPFISAGEAL GIGDILYGTR YAIIDKAAMRN IAPLPABGKF
251 AVIGGLGSAV GFKNWTRAV DRWIQENPNA AETVEAVFNV AAAAKVAKLA
301 KAAKPGKAAV SGDFADSYKK KLALSDSARQ LYQNAKYREA LQKVRISGEI
351 LLG*

```

5 由插入产生的其它氨基酸下加下划线。

总之，730-C1 形式的细胞内表达产生大很高水平的量蛋白质，且对宿主细胞无毒，而天然 C-末端的存在则是有毒的。这些数据表明 730 的“细胞内毒性”与蛋白质的 C-末端 65 氨基酸相关。

10 用有或无前导肽(氨基酸 1-26；缺失得到胞质表达)和有或无 His-标记的表
达，已进行了 ORF29 到前 231 或 368 氨基酸的相等的截短。

实施例 22-961 中的结构域

如上实施例 9 所述，在大肠杆菌中最佳地表达 961 的 GSt 融合。为了改进表达，将蛋白质分成结构域(图 12)。

15 以 YadA(由耶尔森氏菌产生的粘附素，已证明其是位于细胞表面上的粘附素，其形成产生表面凸起的低聚物[Hoiczky 等 (2000)EMBO J 19:5989-99])为基础，设计了 916 的结构域，它们为：前导肽、头部结构域、卷曲-螺旋区域(茎部)和膜锚着点结构域。

20 有或无前导肽表达这些结构域，且任选地融合于 C-末端 His-标记或 N-末端 GST。用 SDS-PAGE 和 Western 印迹，从过夜(o/n)培育或用 IPTG 诱导 3 小时后，分析表达 961 不同结构域的大肠杆菌克隆用于表达的蛋白质的产生和定位。结果如下：

	总裂解物 (Western 印迹)	周质 (Western 印迹)	上清液 (Western 印迹)	OMV SDS-PAGE
961 (o/n)	-	-	-	
961 (IPTG)	+/-	-	-	
961-L (o/n)	+	-	-	+
961-L (IPTG)	+	-	-	+
961c-L (o/n)	-	-	-	
961c-L (IPTG)	+	+	+	
961 Δ_1 -L (o/n)	-	-	-	
961 Δ_1 -L (IPTG)	+	-	-	+

这些结果在大肠杆菌中显示：

- 961-L 是高度表达的且位于外膜上。用 Western 印迹分析已检测到 2 个特异性条带：一条为~45kDa(预计的分子量)，另一条为~180kDa，表明 961-L 可以形成低聚物。另外，这些聚集体在过夜培养物(无 IPTG 诱导)中表达得更多。用该克隆的 OMV 制备物免疫接种小鼠，并获得血清。用过夜培养物(主要为寡聚物形式)的血清是杀菌的；IPTG-诱导的培养物(主要为单体)不是杀菌的。

- 961 Δ_1 -L(在锚区域中部分缺失)是高度表达的，且位于外膜，但不形成低聚物；

- 961c-L(无锚区域)以可溶的形式产生，并被在上清液中输出。

使用 His-融合的 ELISA 和血清杀菌试验中的滴定度如下：

	ELISA	杀菌
961a (aa 24-268)	24397	4096

961b (aa 269-405)	7763	64
961c-L	29770	8192
961c (2996)	30774	>65536
961c (MC58)	33437	16384
961d	26069	>65536

- 用表达不同形式 961(961、961-L、961 Δ_1 -L 和 961c-L)的大肠杆菌克隆研究了 961 是否是粘附素(参见 YadA)。用(a)人上皮细胞和(b)大肠杆菌克隆(过夜培育或 IPTG 诱导 3 小时后)，进行粘附试验。过夜培育的 961-L(961 Δ_1 -L)和 IPTG 诱导的 961c-L(在表面上表达蛋白质的克隆)粘附于人上皮细胞。

也将 961c 用于杂交蛋白(见上)。由于 961 和其结构域变体指导高效表达，

它们理想地适用于杂交蛋白的 N-末端部分。

实施例 23-其它杂交体

以下显示了本发明的其它杂交蛋白(同样参见图 14)。当与单独蛋白质相比
5 时,它们是有利的:

ORF46.1-741

```

1   ATGTCAGATT TGGCAAACGA TTCTTTTATC CGGCAGGTTT TCGACCGTCA
51  GCATTTTCGAA CCGGACGGGA AATACCACCT ATTGGGCAGC AGGGGGGAAC
101 TTGCCGAGCG CAGCGGCCAT ATCGGATTGG GAAAAATACA AAGCCATCAG
151 TTGGGCAACC TGATGATTCA ACAGGCGGCC ATTAAAGGAA ATATCGGCTA
201 CATTGTCCGC TTTTCCGATC ACGGGCACGA AGTCCATTCC CCTTCGACA
251 ACCATGCCTC ACATTCGGAT TCTGATGAAG CCGGTAGTCC CGTTGACGGA
301 TTTAGCCTTT ACCGCATCCA TTGGGACGGA TACGAACACC ATCCCGCCGA
351 CGGCTATGAC GGGCCACAGG GCGGCGGCTA TCCCGCTCCC AAAGCGCGGA
401 GGGATATATA CAGCTACGAC ATAAAGGCG TTGCCAAAA TATCCGCCTC
451 AACCTGACCG ACAACCGCAG CACCGACAA CGGCTTGCCG ACCGTTTCCA
501 CAATGCCGGT AGTATGCTGA CGCAAGGAGT AGGCGACGGA TTCAAACGCG
551 CCACCGGATA CAGCCCGGAG CTGGACAGAT CGGGCAATGC CGCCGAAGCC
601 TTCAACGGCA CTGCAGATAT CGTTAAAAAC ATCATCGGCG CGGCAGGAGA
651 AATTGTCCGC GCAGGCGATG CCGTGCAGGG CATAAAGCGAA GGCTCAAACA
701 TTGCTGTCTT GCAGGGCTTG GGTCTGCTTT CCACCGAAAA CAAGATGGCG
751 CGCATCAACG ATTTGGCAGA TATGGCGCAA CTCAAAGACT ATGCCGACGC
801 AGCCATCCGC GATTGGGCAG TCCAAAACCC CAATGCCGCA CAAGGCATAG
851 AAGCCGTCAG CAATATCTTT ATGGCAGCCA TCCCATCAA AGGGATTGGA
901 GCTGTTCCGG GAAAAACGG CTTGGGCGGC ATCAGGCGAC ATCCTATCAA
951 GCGGTCCGAG ATGGGCGCGA TCGCATTGCC GAAAGGGAAA TCCCGCGTCA
1001 GCGACAATTT TGCCGATGCG GCATACGCCA AATACCCGTC CCTTTACCAT
1051 TCCCGAAATA TCCGTTCAA CTTGGAGCAG CGTTACGGCA AAGAAAACAT
1101 CACCTCCTCA ACCGTGCCGC CGTCAAACGG CAAAATGTC AAAGTGGCAG
1151 ACCAACGCCA CCCGAAGACA GCGGTACCGT TTGACGGTAA AGGGTTTCCG
1201 AATTTTGAGA AGCAGGTGAA ATATGATACG GGATCCGGAG GGGGTGGTGT
1251 CGCCGCCGAC ATCGGTGCGG GGCTTGCCGA TGCACTAACC GCACCGCTCG
1301 ACCATAAAGA CAAAGGTTTG CAGTCTTTGA CGCTGGATCA GTCCGTCAGG
1351 AAAACGAGA AACTGAAGCT GCGGCACAA GGTGCGGAAA AAAGTTATGG
1401 AAACGGTGAC AGCCTCAATA CGGGCAATTT GAAGAACGAC AAGGTCAGCC
1451 GTTTCGACTT TATCCGCCAA ATCGAAGTGG ACGGGCAGCT CATTACCTTG
1501 GAGAGTGGAG AGTTCCAAGT ATACAAACAA AGCCATTCCG CCTTAACCGC
1551 CTTCAGACC GAGCAAATAC AAGATTCCGA GCATTCCGGG AAGATGGTTG
1601 CGAAACGCCA GTTCAGAATC GCGACATAG CGGCGAACA TACATCTTTT

```

1651	GACAAGCTTC	CCGAAAGCGG	CAGGCGGACA	TATCGCGGGA	CGGCGTTCGG
1701	TTCAGACGAT	GCCTGGCGAA	AAC TGACCTA	CACCATAGAT	TTCCCGGCCA
1751	AGCAGGGAAA	CGGCAAAATC	GAACATTTGA	AATCGCCAGA	ACTCAATGTC
1801	GACCTGGCCG	CCGCCGATAT	CAAGCCGGAT	GGAAAACGCC	ATGCCGTCAT
1851	CAGCGGTTC	GTCCTTTACA	ACCAAGCCGA	GAAAGGCAGT	TACTCCCTCG
1901	GTATCTTTGG	CGGAAAAGCC	CAGGAAGTTG	CCGGCAGCGC	GGAAAGTGAA
1951	ACCGTAAACG	GCATACGCCA	TATCGGCCTT	GCCGCCAAGC	AACTCGAGCA
2001	CCACCACCAC	CACCACTGA			

1	MSDLANDSFI	RQVLDRQHFE	PDGKYHLFGS	RGELAERSGH	IGLGKIQSHQ
51	LGNLMIQQA	IKGNIGYIVR	FSDHGHEVHS	PFDNHASHSD	SDEAGSPVDG
101	FSLYRIHWDG	YEHHPADGYD	GPQGGGYAP	KGARDIYSYD	IKGVAQNIRL
151	NLTDRNSTGQ	RLADRFHNAG	SMLTQGVGDG	FKRATRYSPB	LDRSGMAAEA
201	FNGTADIVKN	IIGAAGEIVG	AGDAVQGISB	GSNIAVMHGL	GLLSTENKMA
251	RINDLADMAQ	LKDYAAAAIR	DWAVQNPNA	QGIBAVSNIF	MAAIPKIGIG
301	AVRGKYGLGG	ITAHPIKRSQ	MGAIALPKGK	SAVSDNFADA	AYAKYPSFYH
351	SRNIRSNLEQ	RYGKENITSS	TVPPSNQKNV	KLADQRHPKT	GVPPDGKGFP
401	NFEKHVKYDT	GSGGGGVAAD	IGAGLADALT	APLDHKDKGL	QSLTLDQSVR
451	KNEKLKLAAG	GAKTYGNGD	SLNTGKLNKD	KVSRFDFIRQ	IEVDGQLITL
501	ESGEFQVYKQ	SHSALTAFQT	EQIQDSEHSG	KMVAKRQFRI	GDIAGEHTSF
551	DKLPEGGRAT	YRGTAFGSDD	AGGKLTYYTID	FAAKQGGNGKI	BEHKSPELNV
601	DLAAADIKPD	GKRHAVISGS	VLYNQAEKGS	YSLGIFGGKA	QEVAGSAEVK
651	TVNGIRHIGL	AAKQLEHHHH	HH*		

ORF46.1-961

1	ATGTCAGATT	TGGCAAACGA	TTCTTTTATC	CGGCAGGTTT	TCGACCGTCA
51	GCATTTCGAA	CCCAGCGGGA	AATACCACCT	ATTCCGGCAGC	AGGGGGGAAC
101	TTGCCGAGCG	CAGCGGCCAT	ATCGGATTGG	GAAAAATACA	AAGCCATCAG
151	TTGGGCAACC	TGATGATTCA	ACAGGCGGCC	ATTAAAGGAA	ATATCGGCTA
201	CATTGTCCGC	TTTTCCGATC	ACGGGCACGA	AGTCCATTCC	CCCTTCGACA
251	ACCATGCCTC	ACATTCCGAT	TCTGATGAAG	CCGGTAGTCC	COTTGACCGA
301	TTTAGCCTTT	ACCGCATCCA	TTGGGACGGA	TACGAACACC	ATCCCGCCGA
351	CGGCTATGAC	GGGCCACAGG	GCGGCGGCTA	TCCCCTCTCC	AAAGGCGCGA
401	GGGATATATA	CAGCTACGAC	ATAAAAGGCG	TTGCCCAAAA	TATCCGCCTC
451	AACCTGACCG	ACAACCGCAG	CACCGGACAA	CGGCTTGCCG	ACCGTTTCCA
501	CAATGCCGGT	AGTATGCTGA	CGCAAGGAGT	AGGCGACGGA	TTCAAACGCG
551	CCACCGGATA	CAGCCCCGAG	CTGGACAGAT	CGGGCAATGC	CGCCGAAGCC
601	TTCAACGGCA	CTGCAGATAT	CGTTAAAAAC	ATCATCGGCG	CGGCAGGAGA
651	AATTGTCCGC	GCAGGCGATG	CCGTGCAGGG	CATAAGCGAA	GGCTCAAACA
701	TTGCTGTCTAT	GCACGGCTTG	GGTCTGCTTT	CCACCGAAAA	CAAGATGGCG
751	CGCATCAACG	ATTTGGCAGA	TATGGCGCAA	CTCAAAGACT	ATGCCGCAGC
801	AGCCATCCGC	GATTGGGCRG	TCCAAAACCC	CAATGCCGCA	CAAGGCATAG
851	AAGCCGTCAG	CAATATCTTT	ATGGCAGCCA	TCCCCATCAA	AGGGATTGGA
901	GCTGTTCGGG	GAAAAATACG	CTTGGGCGGC	ATCAGGCGAC	ATCCTATCAA
951	GCGGTCCGAG	ATGGGCGCGA	TCCGATGCC	GAAAGGGAAA	TCCGCGGTCA
1001	GCGACAATTT	TGCCGATGCG	GCATACGCCA	AATACCCGTC	CCCTTACCAT
1051	TCCCGAAATA	TCCGTTCAAA	CTTGGAGCAG	CGTTACGGCA	AAGAAACAT
1101	CACCTCTCTA	ACCGTGCCGC	CGTCAAACGG	CAAAAATGTC	AAACTGGCAG
1151	ACCAACGCCA	CCCGAAGACA	GGCGTACCGT	TTGACGGTAA	AGGGTTTCCG
1201	AATTTTGAGA	AGCACGTGAA	ATATGATACG	GGATCCGGAG	GAGGAGGAGC
1251	CACAAACGAC	GACGATGTTA	AAAAAGCTGC	CACTGTGGCC	ATTGCTGCTG
1301	CCTACAACAA	TGGCCAAGAA	ATCAACGOTT	TCAAAGCTGG	AGAGACCATC
1351	TACGACATTG	ATGAAGACGG	CACAATTACC	AAAAAGACG	CAACTGCAGC
1401	CGATGTTGAA	GCCGACGACT	TAAAGGTCT	GGGTCTGAAA	AAAGTCGTGA
1451	CTAACCTGAC	CAAAACCGTC	AATGAAAACA	AACAAAACGT	CGATGCCAAA
1501	GTAAGAAGCTG	CAGAACTCTGA	AATAGAAAAG	TTAACCAACCA	AGTTAGCAGA
1551	CACGTGATGCC	GCTTTAGCAG	ATACTGATGC	CGCTCTGGAT	GCAACCACCA
1601	ACGCGTTGAA	TAAATTGGGA	GAAAAATATA	CGACATTTGC	TGAAGAGACT
1651	AAGACAATA	TCGTAAAAAT	TGATGAAAAA	TTAGAAGCCG	TGGCTGATAC
1701	CGTCGACAAG	CATGCCGAAG	CATTCAACGA	TATCGCCGAT	TCATTGGATG
1751	AAACCAACAC	TAAGGCAGAC	GAAGCCGTC	AAACCGCCAA	TGAAGCCAAA
1801	CAGACGGCCG	AAGAAACCAA	ACAAAACGTC	GATGCCAAAG	TAAAGCTGTC
1851	AGAAACTGCA	GCAGGCAAAAG	CCGAAGCTGC	CGCTGGCACA	GCTAATACTG
1901	CAGCCGACAA	GGCCGAAGCT	GTCGCTGCAA	AAGTTACCGA	CATCAAAGCT
1951	GATATCGCTA	CGAACAAAGA	TAATATTGCT	AAAAAAGCAA	ACAGTGCCGA
2001	CGTGTACACC	AGAGAAGAGT	CTGACAGCAA	ATTTGTGAGA	ATTGATGGTC

```

2051 TGAACGCTAC TACCGAAAA TTGGACACAC GCTTGGCTTC TGCTGAAAA
2101 TCCATTGCCG ATCAGGATAC TCGCCTGAAC GGTTTGGATA AAACAGTGT
2151 AGACCTGCCG AAAGAAACCC GCCAAGGCCT TGCAGAACAA GCCGCGCTCT
2201 CCGGTCTGTT CCAACCTTAC AACGTGGGTC GGTTCATGT AACGGCTGCA
2251 GTCGGCGGCT ACAAATCCGA ATCGGCAGTC GCCATCGGTA CCGGCTTCCG
2301 CTTTACCGAA AACTTTGCCG CCAAAGCAGG CGTGGCAGTC GGCACTTCGT
2351 CCGGTTCTTC CGCAGCCTAC CATGTCGGCG TCAATTACGA GTGGCTCGAG
2401 CACCACCACC ACCACCCTG A

```

```

1 MSDLANDSFI RQVLDRQHFE PDGKYHLFGS RGLAERSGH IGLGKIQSHQ
51 LGNLMIQQAA IKGNIGYIVR FSDHGHEVHS PFDNHASHSD SDEAGSPVDG
101 FSLYRIHWDG YEHHPADGYD GPQGGGYPAP KGARDIYSYD IKGVAQNIRL
151 NLTDNRSTGQ RLADRFHAG SMLTQGVGDG FKRATRYSPF LDRSGNAABA
201 FNGTADIVKN IIGAAGEIVG AGDAVQGIS ESNLAVMHGL GLLSTENKMA
251 RINDLADMAQ LKDYAAAAIR DWAVQMFNAA QGIEAVSNIF MAAIPKIGIG
301 AVRKYGLGG ITAHPIKRSQ MGAIALPKGK SAVSDMFADA AYAKYPSFYH
351 SRNIRSNLEQ RYKKNITSS TVPPSNKKNV KLADQRHPKT GVPFDGKGFP
401 NFEKHVKYDT GSGGGGATND DDVKAATVA LAAAYNNGQE INGPKAGETI
451 YDIDEDGTIT KRDATAADVE ADDFKGLGLK KVVTLNLTIV NENKQNVDAK
501 VKAAESEIEK LTTKLADTDA ALADTDAALD ATTNALNKLG ENITTFABET
551 KTNIVKIDEX LEAVADTVDK HAAAFNDIAD SLDEFINTKAD EAVKTANEAK
601 QTAETKQNV DAKVKAETA AGKABAAAGT ANTAADKAEA VAAKVTDIKA
651 NPLNNDNIA KANSADVYT REESDSKFVR IDGLNATTEK LDTRLASAEK
701 SLADHDTRLN GLDRTVSDLR KETROGLAEQ AALSGLPQPY NVGRFNVTA
751 VGGYKSES AV AIGTGFRFTE NFAAKAGVAV GTSSGSSAAY HVGVNVEWLE
801 HHHHHH*

```

ORF46.1-961g

```

1 ATGTCAGATT TGGCAAACGA TTCTTTTATC CGGCAGGTTT TCGACCGTCA
51 GCATTTCGAA CCCGACGGGA AATACCACCT ATTCGGCAGC AGGGGGGAAC
101 TTGCCGAGCG CAGCGGCCAT ATCGGATTGG GAAAAATACA AAGCCATCAG
151 TTGGGCAACC TGATGATTCA ACAGCCGGCC ATTAAGGAA ATATCGGCTA
201 CATTGTCCGC TTTTCCGATC ACGGACACGA AGTCCATCC CCTTCGACA
251 ACCATGCCCTC ACATTCCGAT TCTGATGAAG CCGGTAGTCC CGTTGACCGA
301 TTAGCCTTTT ACCGCATCCA TTGGGACCGA TACGAACACC ATCCCGCGCA
351 GCGCTATGAC GGGCCACAGG GCGGCGGCTA TCCCGCTCCC AAAGGCGCGA
401 GGGATATATA CAGCTACGAC ATAAAAGGCG TTGCCCAAAA TATCCGCTC
451 AACCTGACCG ACAACCGCAG CACCGACAA CGGCTTCCG ACCGTTTCCA
501 CAATGCCGGT AGTATGCTGA CGCAAGGAGT AGGCGACCGA TTCAAACGCG
551 CCACCCGATA CAGCCCCGAG CTGGACAGAT CGGGCAATGC CGCCGAAGCC
601 TTCAAACGCG CTGCAGATAT CGTTAAAAAC ATCATCGGCG CGGCAGGAGA
651 AATTGTCCGC GCAGGCGATG CCGTGCAGGG CATAAGCGAA GGCTCAACA
701 TTGCTGTCTAT GCACGGCTTG GGTCTGCTTT CCACCGAAAA CAAGATGGCG
751 CGCATCAACG ATTTGGCAGA TATGGCGCAA CTCAAAGACT ATGCCGAGC
801 AGCCATCCGC GATTGGGCGA TCCAAAACCC CAATGCCGCA CAAGGCATAG
851 AAGCCGTCAG CAATATCTTT ATGGCAGCCA TCCCATCAA AGGGATTGGA
901 GCTGTTCGGG GAAAATACCG CTTGGGCGGC ATCAGGCAC ATCTATCAA
951 GCGGTCCGAG ATGGGCGCGA TCGCATTGCC GAAAGGGAAA TCCGCGTCA
1001 GCGACAATTT TGCCGATGCG GCATACGCCA AATACCCGTC CCTTACCAT
1051 TCCCGAATA TCCGTTCAA CTTGGAGCAG CGTTACGCGA AAGAAAACAT
1101 CACCTCCTCA ACCGTGCCGC CGTCAAACGG CAAAAATGTC AAAGTGGCAG
1151 ACCAACGCCA CCCGAAGACA GCGTACCCTT TGACCGTAA AGGGTTTCCG
1201 AATTTTGAGA AGCACGTGAA ATATGATACG GGATCCGAG GAGGAGGAGC
1251 CACAAACGAC GACGATGTTA AAAAAGCTGC CACTGTGGCC ATTGCTGCTG
1301 CCTACACAA TGGCCAAGAA ATCAACGGTT TCAAAGCTGG AGAGACCATC
1351 TACGACATTG ATGAAGACGG CACAATTACC AAAAAAGACG CAACTGCAGC
1401 CGATGTTGAA GCCGACGACT TTAAGGCTCT GGGTCTGAAA AAGTCTGTA
1451 CTAACTGAC CAAAACCGTC AATGAAAACA AACAAAACGT CGATGCCAAA
1501 GTAAAAGCTG CAGAACTGTA AATAGAAAAG TTAACAACCA AGTTAGCAGA
1551 CACTGATGCC GCTTTAGCAG ATACTGATGC CGCTCTGGAT GCAACCACCA
1601 ACGCCTTGAA TAAATTGGGA GAAAAATATA CGACATTTGC TGAAGGACT
1651 AAGACAAATA TCGTAAAAAT TGATGAAAAA TTAGAAGCCG TGGCTGATAC
1701 CGTCGACAA GATGCCGAAG CATTCAACGA TATCGCCGAT TCATTGGATG
1751 AAACCAACAC TAAGGCAGAC GAAGCCGTCA AAACCGCCAA TGAAGCCAAA
1801 CAGACGGCCG AAGAAACCAA ACAAACGCTC GATGCCAAG TAAAAGCTGC
1851 AGAAACTGCA GCAGGCAAG CCGAAGCTGC CGCTGGCACA GCTAATCTG

```

```

1901 CAGCCGACAA GGCCGAAGCT GTCGCTGCAA AAGTTACCGA CATCAAAGCT
1951 GATATCGCTA CGAACAAAGA TAATATTGCT AAAAAAGCAA ACAGTGCCGA
2001 CGTGTACACC AGAGAAGAGT CTGACAGCAA ATTTGTGAGA ATTGATGGTC
2051 TGAACGCTAC TACCGAAAAA TTGGACACAC GCTTGGCTTC TGCTGAAAAA
2101 TCCATTGCCG ATCAGCATAC TCGCCTGAAC GGTTTGGATA AAACAGTGTC
2151 AGACCTGCGC AAAGAAACCC GCCAAGGCCT TGCAGAACAA GCCGCGCTCT
2201 CCGGTCTGTT CCAACCTTAC AACGTGGGTC TCGAGACCAA CCACCACCAC
2251 CACTGA

```

```

1 MSDLANDSFI RQVLDRQHFE PDGKYHLFGS RGELAERSGH IGLGKIQSHQ
51 LGNLMIQQAA IKGNIGYIVR FSDHGHEVHS PFDNHASHSD SDEAGSPVDG
101 FSLYRIHWDG YEHHPADGYD GPQGGGYFAP KGARDIYSYD IKGVAQNIIRL
151 NLTDNRSTGQ RLADRFHNAG SMLTQGVGDG FKRATRYSPZ LDRSGNAABA
201 FNGTADIVKN IIGAAGEIVG AGDAVQGISE GSNIAVMHGL GLLSTENKMA
251 RINDLADMAQ LKDYAAAAIR DWAQVQNFNA QGIEAVSNIF MAALPIKQIG
301 AVRKGKYLGG ITAHPIKRSQ MGAIALPKGK SAVSDNFADA AYAKYPSFYH
351 SRNIRSNLEQ RYKKNITSS TVPPSNGKNV KLADQRHPKT GVPFDGKGFY
401 NFEKHVKYDT GSGGGGATND DDVKAATVA IAAAYNNGQE INGFKAGETI
451 YDIDEDGTIT KKDATAADVE ADDFKGLGLK KVVNLTITKV MENKQNVDAK
501 VKAAESEIEK LITKLADTDA ALADTDAALD ATTNALNKLK ENITTFAKET
551 KTNIVKIDDK LEAVADTVDK HAEAFNDIAD SLDETNTKAD EAVKTANEAK
601 QTAETKQNV DAKVKAAETA AGKAEAAAGT ANTAADKAEA VAAKVTDIKA
651 DIATNKDNIA KKANSADVYT REESDSKFVR IDGLNATTEK LDTRLASAEK
701 SIADHDTRLN GLDKTVSDLR KETROGLAEQ AALSGLFQPY NVGLEHHHHE
751 H*

```

961-ORF46.1

```

1 ATGGCCACAA ACGACGACGA TGTAAAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCCTAC AACAAATGGCC AAGAAATCAA CGGTTCCTAAA GCTGGAGAGA
101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
201 CGTGCATAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAAATAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCAAT
501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
551 CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACCTCGATGC CAAAGTAAAA
601 GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
651 TACTGCAGCC GACAAGGCCG AAGCTGTGCG TGCAAAAGTT ACCGACATCA
701 AAGCTGATAT CGCTACGAAC AAGATAATA TTGCTAAAAA AGCAAAACAT
751 GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAAATTT TCAGAAATGA
801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
851 AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
901 GTGTCAGACC TGCACAAAGA AACCCGCCAA GGCCCTGCGA AACAGCCCG
951 GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTCGGTTC AATGTAACGG
1001 CTGCAGTCGG CGGCTACAAA TCCGAATCGG CAGTCGCCAT CGGTACCGGC
1051 TTCCGCTTTA CCGAAAACCT TGCCGCCAAA GCAGGCGTGG CAGTCGGCAC
1101 TTCTCCGGT TCTTCCGCAG CCTACCATGT CGGCGTCAAT TACGAGTGGG
1151 GATCCGGAGG AGGAGGATCA GATTGGCAA ACGATTCTTT TATCCGGCAG
1201 GTTCTCGACC GTCAGCATTT CGAACCCGAC GGGAAATACC ACCTATTCCG
1251 CAGCAGGGGG GAACTTGCCG AGCGCAGCGG CCATATCGGA TTGGGAAAAA
1301 TACAAAGCCA TCAATTGGGC AACCTGATGA TTCAACAGGC GGCCATTAAA
1351 GGAAATATCG GCTACATTGT CCGCTTTTCC GATCACGGGC ACGAAGTCCA
1401 TTCCCCCTTC GACAACCATG CCTCACATTC CGATTCTGAT GAAGCCGGTA
1451 GTCCCGTTGA CGGATTTAGC CTTTACCGCA TCCATTGGGA CGGATACGAA
1501 CACCATCCCG CCGACGGCTA TGACGGGCCA CAGGGCGGCG GCTATCCCGC
1551 TCCCAAAGGC GCGAGGGATA TATACAGCTA CGACATAAAA GGCGTTGCCG
1601 AAAATATCOG CCTCAACCTG ACCGACAACC GCAGCACCAG ACAACGGCTT
1651 GCGGACCGTT TCCACAATGC CGGTAGTATG CTGACGCAAG GAGTAGGCGA
1701 CGGATTCAAA CGCGCCACCC GATACAGCCC CGAGCTGGAC AGATCGGGCA
1751 ATGCCGCCGA AGCCTTCAAC GGCACGCGAG ATATCGTTAA AAACATCATC
1801 GGCGCGGCAG GAGAAATTGT CGGCGCAGGC GATGCCGCTG AGGGCATAAG
1851 CGAAGGCTCA AACATTGCTG TCATGCAACG CTGGGGTCTG CTTTCCACCG
1901 AAAACAAGAT GGCGCGCATC AACGATTGGG CAGATATGGC GCAACTCAAA

```

```

1951 GACTATGCCG CAGCAGCCAT CCGCGATTGG GCAGTCCAAA ACCCCAATGC
2001 CGCACAAAGGC ATAGAAAGCCG TCAGCAATAT CTTTATGGCA GCCATCCCCA
2051 TCAAAGGGAT TGGAGCTGTT CGGGGAAAAT ACGGCTTGGG CGGCATCACG
2101 GCACATCCTTA TCAAGCGGTC GCAGATGGGC GCGATCGCAT TGCCGAAAGG
2151 GAAATCCGCC GTCAAGCACA ATTTTGCCGA TCGCGCATAC GCCAAATACC
2201 CGTCCCCTTA CCATTCGCCA AATATCCGTT CAAACTTGA GCAGCGTTAC
2251 GGCAAAGAAA ACATCACCTC CTCACCGTG CCGCCGTCAA ACOGCAAAAA
2301 TGTCAAACTG GCAGACCAAC GCCACCCGAA GACAGGCGTA CCGTTTGACG
2351 GTAAAGGGTT TCCGAATTTT GAGAAGCACG TGAATATGA TACGCTCGAG
2401 CACCACCACC ACCACCCTG A

```

```

1 MATNDDVVK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
51 AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDKVKAAE SEIEKLTTKL
101 ADTDAALADT DAALDATTNA LNKLGENTIT FAKETKTNIV KIDEKLEAVA
151 DFTVDKHAFAF NDIADSLDET NTKADRAVKT ANEAKQTAKE TKQNVDAKVK
201 AAETAAGKAE AAAGTAMTAA DKAEAVAAKV TDIKADIATN KDNIAKKANS
251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
301 VSDLRKETRQ GLAEQAALSG LFQPYNVGRF NVTAAVGGYK SESAVAIGTG
351 FRFTENFAAK AGVAVGTSSG SSAAYHVGUN YEWGSGGGGS DLANDSFIRQ
401 VLDRQEFEPD GKYLFGSRG ELAERSGHIG LGKIQSHQLG NLMIQQAIAK
451 GNIGYIVRFS DHGHEVHSFF DNEASHSDSD BAGSPVDGFS LYRIHMDGYE
501 HHPADGYDGP QGGGYPAFKG ARDIYSYDIK GVAQNIRLNL TDNRSTGQRL
551 ADRFENAGSM LTQGVGDGFK RATRYSPELD RSGNAEAFN GTADIVKNII
601 GAAGEIVGAG DAVQGISSEGS NLAVMHGLGL LSTENFMARI NDLDMAQLK
651 DYAAAIRDW AVQNPNAAG IEAVSNIFMA AIPKIGIGAV RGKYGLGGIT
701 AHPKRSQMG AIALPKGKSA VSDNFADAAY AKYPSFYHSR NIRSNLBQRY
751 GKENITSSTV PPSNGKNVKL ADQREPKTGV PFDGKGFPNF EKHVKYDTLE
801 HHHHHH*

```

961-741

```

1 ATGGCCACAA ACGACGACGA TGTAAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCTTAC AACAAATGCC AAGAAATCAA CGGTTTCAA GCTGGAGAGA
101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAA GGTCTGGGTC TGA AAAAAGT
201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 CACCACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCAIT
501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
551 CCAAACGAC GGCCGAAGAA ACCAAACAAA ACCTCGATGC CAAAGTAAAA
601 GCTGCAGAAA CTGCGACGAG CAAGCCGAA GCTGCCGCTG GCACAGCTAA
651 TACTGCAGCC GACAAGGCCG AAGCTGTGCG TGCAAAAGTT ACCGACATCA
701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAAACAGT
751 GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAAATTTG TCAGAAATGA
801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
851 AAAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGOTTT GGATAAAACA
901 GTGTCAAGCC TGCGCAAAGA AACCCGCCAA GGCCTTGCA GACAAGCCGC
951 GCTCTCCGGT CTGTTCCAA CTTACAACGT GGGTCCGTTT AATGTAACGG
1001 CTGCACTCGG CGGCTACAAA TCCGAATCGG CAGTCGCCAT CGGTACCGGC
1051 TTCCGCTTTA CCGAAACTTT TGCCGCCAAA GCAGGCGTGG CAGTCGGCAC
1101 TTGCTCCGGT TCTTCCGCG CTTACCATGT CCGCGTCAAT TACGAGTGGG
1151 GATCCGGAGG GGGTGGTGTG GCCGCCGACA TCGGTGCGGG GCTTGCCGAT
1201 GCATAACCG CACCGCTCGA CCATAAAGAC AAAGTTTTC AGTCTTTGAC
1251 GCTGGATCAG TCCGTCAGGA AAAACGAGAA ACTGAAGCTG GCGGCACAAG
1301 GTGCGGAAAA AACTTATGGA AACGGTGACA GCCTCAATAC GGGCAAATTG
1351 AAGAACGACA AGGTACGCGG TTTGACTTT ATCCGCCAAA TCGAAGTGGA
1401 CGGGCAGCTC ATTACCTTGG AGAGTGAGGA GTTCCAAGTA TACAAACAAA
1451 GCCATCCCG CTTAACCGCC TTTGAGACCG AGCAAAATACA AGATTCGGAG
1501 CATTCGGGGA AGATGGTTGC GAAACGCCAG TTCAGAAATCG GCGACATAGC
1551 GGGCGAACAT ACATCTTTTG ACAAGCTTCC CGAAGGCCGC AGGGCGACAT
1601 ATCGCGGGAC GCGGTTCGGT TCAGACGATG CCGGCGGAAA ACTGACCTAC
1651 ACCATAGATT TCGCCGCCAA GCAGGGAAC GGCAAAATCG AACATTTGAA
1701 ATCGCCAGAA TCAATGTG ACCTGGCCGC CGCCGATATC AAGCCGGATG
1751 GAAAAAGCCA TGCCGTATC AGCGGTTCG TCTTTTACAA CCAAGCCGAG
1801 AAAGGCAGTT ACTCCCTCGG TATCTTTGGC GGAAAAGCCC AGGAAGTTGC

```

1851 CCGCAGCGCG GAAGTGA AAA CCGTAAACGG CATACGCCAT ATCGGCCTTG
1901 CCGCCAAGCA ACTCGAGCAC CACCACCACC ACCACTGA

1 MATNDDDVKK AATVAIAAAY NNGQEINGFK AGETIYDIDE DGTITKKDAT
51 AADVEADDFK GLGLKKVVVN LTKTVNENKQ NVDAKVKAAB SEIEKLTTKL
101 ADTDAALADT DAALDATTNA LNKLGENTIT FAEETKTNTV KIDEKLEAVA
151 DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAEZ TKQNVDAKVK
201 AAETAAGKAE AAAGTANTAA DKABAVAAKV TDIKADIATN KDNIAKKANS
251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
301 VSDLRKETRQ GLAEQAALSG LFQPFYNVGRF NVTAAVGGYK SESAVAIGTG
351 FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEWGSGGGGV AADIGAGLAD
401 ALTAPLDHKD KGLQSLTLQ SVRKNEKLKL AAQGAEKTYG NGDSLBTGKL
451 KNDKVSFDF IRQLVDGQL ITLESSEFQV YKQSHSALTA FQTEQIQDSE
501 HSGKMVAKRQ FRIGDIAGEH TSFDKLFEGG RAYRTGTAFG SDDAGGKLT
551 TIDFAAKQGN GKIEHLKSPB LNVDLAAADI KPDGKRHAVI SSVLYNQAE
601 KGSYSLGIFG GKAEVAGSA EVKTVNGIRH IGLAAKQLEH HHHHH*

961-983

1 ATGGCCACAA ACGACGACGA TGTTAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCTAC AACAAATGGCC AAGAAATCAA CGTTTTCAAA GCTGGAGAGA
101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGA AAAAAGT
201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAA AGCTGCAGAA TCTGAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCATT
501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
551 CCAAACGAGC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
601 GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
651 TACTGCAGCC GACAAGGCCG AAGCTGTGCG TGCAAAAAGTT ACCGACATCA
701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAACAGT
751 GCCGCGTGT ACACCAGAGA AGAGCTGAC AGCAAATTTG TCAGAATTGA
801 TGGTCTGAAC GCTACTACCG AAAAAATTGA CACACGCTTG GCTTCTGCTG
851 AAAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
901 GTGTCTAGACC TGCCGAAAGA AACC CGCCAA GGCTTTCGAG AACAGCCGC
951 GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTCGGTTC AATGTAACGG
1001 CTGCGTCCG CGGCTACAAA TCCGAATCGG CAGTCGCCAT CGGTACCGGC
1051 TTCCGCTTTA CCGAAAACCT TGCCGCCAAA GCAGGCGTGG CAGTCGCCAC
1101 TTCGTCCGGT TCTTCCGCG CCTACCATGT CGGCGTCAAT TACGAGTGGG
1151 GATCCGCGCG AGGCGGCACT TCTGCGCCCG ACTTCAATGC AGGCGGTACC
1201 GGTATCGGCA GCAACAGCAG AGCAACAACA GCGAAATCAG CAGCAGTATC
1251 TTACGCCGGT ATCAAGAACG AAATGTGCA AGACAGAAGC ATGCTCTGTG
1301 CCGGTCCGGA TGACGTTGCG GTTACAGACA GGGATGCCAA AATCAATGCC
1351 CCCCCCCGA ATCTGCATAC CGGAGACTTT CCAAACCCAA ATGACGCATA
1401 CAAGAATTTG ATCAACCTCA AACCTGCAAT TGAAGCAGGC TATACAGGAC
1451 GCGGGGTAGA GGTAGGTATC GTCGACACAG GCGAATCCGT CGGCAGCATA
1501 TCCTTTCCCG AACTGTATGG CAGAAAAGAA CACGGCTATA ACGAAAATTA
1551 CAAAACTAT ACGGCGTATA TGCGGAAGGA AGCGCCTGAA GACGGAGGCG
1601 GTAAAGACAT TGAAGCTTCT TTCGACGATG AGGCCGTTAT AGAGACTGAA
1651 GCAAAGCCGA CGGATATCCG CCACGTAAAA GAAATCGGAC ACATCGATTT
1701 GGTCTCCCAT ATTATGGCG GCGGTTCCGT GGACGGCAGA CCTGCAGCG
1751 GTATTGCGCC CGATGCGACG CTACACATAA TGAATACGAA TGATGAAACC
1801 AAGAACGAAA TGATGGTTGC AGCCATCCGC AATGCATGGG TCAAGCTGGG
1851 CGAACGTGGC GTGCGCATCG TCAATAACAG TTTTGAACA ACATCGAGGG
1901 CAGGCACGTC CGACCTTTTC CAAATAGCCA ATTCCGAGGA GCAGTACCGC
1951 CAAGCGTTGC TCGACTATT CCGCGGTGAT AAAACAGACG AGGGTATCCG
2001 CCTGATGCAA CAGAGCGATT ACGGCAACCT GTCTTACCAC ATCCGTAAATA
2051 AAAACATGCT TTTTATCTTT TCGACAGGCA ATGACGCACA AGCTCAGCCC
2101 AACACATATG CCTATTGCG ATTTTATGAA AAAGACGCTC AAAAAGGCAT
2151 TATCACAGTC GCAGGCGTAG ACCGCACTGG AGAAAAGTTC AAACGGGAAA
2201 TGTATGGAGA ACCGGGTACA GAACCGCTTG AGTATGGCTC CAACCATTGC
2251 GGAATTACTG CCATGTGGTG CCTGTCCGCA CCCTATGAAG CAAGCGTCCG
2301 TTTACCCGT ACAAACCCGA TTCAAATTGC CGGAACATCC TTTTCCGCAC
2351 CCATCGTAAC CGGCACGGCG GCTCTGCTGC TGCAGAAATA CCCGTGGATG

2401 AGCAACGACA ACCTGCGTAC CACGTTGCTG ACGACGGCTC AGGACATCGG
 2451 TGCAGTCCGC GTGGACACCA AGTTGGGCTG GGGACTGCTG GATGCGGGTA
 2501 AGGCCATGAA CGGACCCGCG TCCTTTCCGT TCGGCGACTT TACCGCCGAT
 2551 ACGAAAGGTA CATCCGATAT TGCCCTACTCC TTCCGTAAACG ACATTTCAGG
 2601 CACGGGCGGC CTGATCAAAA AAGGCGGCAG CCAACTGCAA CTGCACGGCA
 2651 ACAACACCTA TACGGGCAAA ACCATTATCG AAGGCGGTTT GCTGGTGTG
 2701 TACGGCAACA ACAAATCGGA TATGCGCGTC GAAACCAAAG GTGCGCTGAT
 2751 TTATAACGGG GCGGCATCCG GCGGCAGCCT GAACAGCGAC GGCATTGTCT
 2801 ATCTGGCAGA TACCGACCAA TCCGGCGCAA ACGAAACCGT ACACATCAAA
 2851 GGCAGTCTGC AGCTGGACGG CAAAGGTACG CTGTACACAC GTTTGGGCAA
 2901 ACTGCTGAAA GTGGACGGTA CGCGGATTAT CGGCGGCAAG CTGTACATGT
 2951 CGGCACGGCG CAAGGGGGCA GGCTATCTCA ACAGTACCGG ACGACGTGTT
 3001 CCCTTCCTGA GTGCCGCCAA AATCGGGCAG GATTATTTCT TCTTCACAAA
 3051 CATCGAAACC GACGGCGGCC TGCTGGCTTC CCTCGACAGC GTCGAAAAAA
 3101 CAGCGGGCAG TGAAGGCGAC ACCTGTCTCT ATTATGTCCG TCGCGGCAAT
 3151 GCGGCACGGA CTGCTTCGCG AGCGGCACAT TCCGCGCCCG CGGCTCTGAA
 3201 ACACGCCGTA GAACAGGGCG GCAGCAATCT GGAAACCTG ATGCTCGAAC
 3251 TGGATGCCTC CGAATCATCC GCAACACCCG AGACGGTTGA AACTGCGGCA
 3301 GCCGACCGCA CAGATATGCC GGGCATCCGC CCCTACGGCG CAACTTTCCG
 3351 CGCAGCGGCA GCCGTACAGC ATGCGAATGC CGCCGACGGT GTACGCATCT
 3401 TCACAGTCT GTGCTTACC GTCTATGCCG ACAGTACCGT CGCCATGCCC
 3451 GATATGCAGG GACGCCGCTT GAAAGCCGTA TCGGACGGGT TGGACCACAA
 3501 CGGCACGGGT CTGCGCGTCA TCGCGCAAAC CCAACAGGAC GGTGGAACGT
 3551 GGGAAACAGG CGGTGTGTA GCGAAAATGC GCGGCAGTAC CCAAACCGTC
 3601 GGCATTGCCG CGAAAACCGG CGAAAATACG ACAGCAGCCG CCACACTGGG
 3651 CATGGGACGC AGCACATGGA GCGAAAACAG TGCAAATGCA AAAACCGACA
 3701 GCATTAGTCT GTTTGCAGGC ATACGGCACG ATGCGGGCGA TATCGGCTAT
 3751 CTCAAAGGCC TGTTCCTCTA CGGACGCTAC AAAACAGCA TCAGCCGCG
 3801 CACCGGTGCG GACGAACATG CGGAAGGCAG CGTCAACGGC ACGCTGATGC
 3851 AGCTGGGCGG ACTGGGCGGT GTCAACGTTT CGTTTGCCGC AACGGGAGAT
 3901 TTTGACGGTCG AAGCGGCTCT GCGCTACGAC CTGCTCAAAT AGGATGCATT
 3951 CGCCGAAAAA GGCAGTGTCT TGGCTGGAG CGGCAACAGC CTCACTGAAG
 4001 GCACGCTGGT CGGACTCGCG GGTCTGAAGC TGTCGCAACC CTTGAGCGAT
 4051 AAACCCGTCC TGTTTGCAAC GCGCGGCGTG GAACCGGACC TGAACGGACG
 4101 CGACTACACG GTAACGGGCG GCTTTACCGG CGGCACTGCA GCAACCGGCA
 4151 AGACGGGGCG ACGCAATATG CCGCACACCC GTCTGGTTGC CGGCTGGGCG
 4201 GCGGATGTGC AATTCGGCAA CGGCTGGAAC GGCTTGGCAC GTTACAGCTA
 4251 CGCCGGTTCC AAACAGTACG GCAACCACAG CGGACGAGTC GGCCTAGGCT
 4301 ACCGGTTTCT CGAGCACCAC CACCACCACC ACTGA

1 MATNDDVVK AATVALAAAY NNGQELNGFK AGETIYDIDE DGTITKEDAT
 51 AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAEE SEIEKLTTKL
 101 ADTDAALADT DAALDATTA LNKLGENTTT FAEETKTNTV KIDEKLEAVA
 151 DTVDKHABAF NDIADSLDET NTKADEAVKT ANEAKQTAE TKQNVDAKVK
 201 AAETAAGKAE AAAGTANTAA DKAEVAAKV TDIKADIATN KDNIAKKANS
 251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSLADH DTRLNGLDKT
 301 VSDLRKETRQ GLAEQAALSG LFQFYNVGRF NVTAAVGGYK SESAVAIGTG
 351 FRFTENFAAK AGVAVGTSSG SSAAYHVGVN YEWGSGGGGT SAPDFNAGGT
 401 GIGSNRATT AKSAAVSYAG IKNEMCKDRS MLCAGRDDVA VTDRDAKINA
 451 PPPNLHTGDF PNPNDAYKNL INLKPAIBAG YTGRGVEVGI VDTGESVSGI
 501 SFPELYGRKE HGYNENYKNY TAYMRKEAPE DGGGKDIEAS FDDEAVIETE
 551 AKPTDIRHVK EIGHIDLVS HIGGRSVDGR PAGGIAPDAT LHIMNTDET
 601 KNEMVAAIR NAWVKLGERG VRIVNNSFGT TSTRAGTADLF QIANSEBQYR
 651 QALLDYSGGD KTDEGIRLMQ QSDYGNLSYH IRNKNMLFIF STGNDAAQAP
 701 NTYALLPFYE KDAQEGITV AGVDRSGEKF KREMYGEPGT EPLEYGSNHC
 751 GITAMNCLSA FYEASVRFTR TNPIQIAGTS FSAPIVTGTA ALLLQKYPWM
 801 SNDNLRTLL TTAQDIGAVG VDSKFGWGLL DAGKAMNGPA SFFFGDFTAD
 851 TKGTSDIAYS FRNDISGTGG LIKKGGSQLO LHGNNTYTGK TIEGGSLVL
 901 YGNKSDMRV ETKGALIYNG AASGGSLSND GIVYLADTDQ SGANETVHIK
 951 GSLQLDGKGT LYTRLGLKLLK VDGTALIGK LYMSARGKGA GYLNSTGRRV
 1001 PFLSAARIQ DYSPFTNIET DGGLLASLDS VEKTASSEGD TLYYVVRGN
 1051 AARTASAAAH SAPAGLKHAV EQGGSNLENL MVELDASESS ATPETVETAA
 1101 ADRTDMPGIR PYGATFRAAA AVQHANAADG VRIFNSLAAT VYADSTAAHA
 1151 DMQGRRLKAV SDGLDENGTO LRVLAQTQOD GGTWEQGGVE GKMRGSGTQV
 1201 GIAAKTGENT TAAATLGMGR STWSENSANA KTDISISLFAG IREDAGDIGY
 1251 LKGLFSYORY KNSISRSTGA DEHAROSVNG TLMQLGALGG VNVFFAATGD
 1301 LTVEGGLRYD LKQDAFAEK GSALGWSGNS LTBGTLVGLA GLKLSQPLSD

1351 KAVLFATAGV ERDLNGRDYT VTGGFTGATA ATGKTGARNM PETRLVAGLG
1401 ADVEFGNGWN GLARYSYAGS RQYGNHSGRV GVGYRFLHH HHHH*

961g-ORF46.1

1 ATGGCCACAA ACGACGACGA TGTTAAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCCTAC AACAAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
101 CCATCTACGA CATGTATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGA AAAAAGT
201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 ACCCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTCAAT
501 GGATGAAACC AACACTAAGG CAGACGAGC CGTCAAAACC GCCAATGAAG
551 CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
601 GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
651 TACTGCAGCC GACAAGGCCG AAGCTGTGCG TGCAAAAGTT ACCGACATCA
701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
751 GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAAATTTG TCAGAAATGA
801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
851 AAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAAACA
901 GTGTCAGACC TGCGCAAGAA AACCCGCCAA GGCTTGCAG AACAAAGCCG
951 GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTGGATCC GGAGGAGGAG
1001 GATCAGATTT GGCAACGAT TCTTTTATCC GGCAGGTTCT CGACCGTCAG
1051 CATTTGCAAC CCGACGGGAA ATACCACCTA TTCGGCAGCA GGGGGGAAGT
1101 TGCCGAGCGC AGCGGCCATA TCGGATTGGG AAAAATACAA AGCCATCAGT
1151 TGGCAACCT GATGATTCAA CAGGCGGCCA TTAAAGGAAA TATCGGCTAC
1201 ATTGTCCGCT TTTCCGATCA CGGGCACGAA GTCCATTCCC CCTTCGACAA
1251 CCATGCCTCA CATTCGATTT CTGATGAAGC CGGTAGTCCC GTTGACGGAT
1301 TTAGCCTTTA CCGCATCCAT TGGGACGGAT ACGAACACCA TCCCGCCGAC
1351 GGCTATGACG GGCCACAGGG CGGCGGCTAT CCCGCTCCA AAGGCGCGAG
1401 GGATATATAC AGCTACGACA TAAAGGCGT TGCCCAAAAT ATCCGCCTCA
1451 ACCTGACCGA CAACCGCAGC ACCGACAAAC GGCTTGCCGA CCGTTTCCAC
1501 AATGCCGGTA GTATGCTGAC GCAAGGAGTA GGCAGACGAT TCAAAACGCG
1551 CACCCGATAC AGCCCGGAGC TGGACAGATC GGGCAATGCC GCCGAAGCCT
1601 TCAACGGCAG TGCAGATATC GTTAAAAACA TCATCGGCGC GGCAGGAGAA
1651 ATTGTCCGGC CAGGCGATGC CGTGCAGGGC ATAAGCGAAG GCTCAAACAT
1701 TGCTGTGATG CACGGCTTGG GTCTGCTTTC CACCGAAAAC AAGATGGCGC
1751 GCATCAACGA TTTGGCAGAT ATGGCGCAAC TCAAAGACTA TGCCGACGCA
1801 GCCATCCGCG ATTGGGCAAT CCAAAACCCC AATGCCGCAC AAGGCATAGA
1851 AGCCGTGAGC AATATCTTTA TGGCAGCCAT CCCCATCAA GGGATTGGAG
1901 CTGTTCCGGG AAAATACGGC TTGGGCGGCA TCACGGCACA TCCTATCAAG
1951 CGGTGCGAGA TGGGCGCGAT CGCATTGCCG AAAGGGAAAT CCGCCGTCAG
2001 CGACAAATTT GCCGATGCGG CATACGCCAA ATACCCGTC CCTTACCATT
2051 CCCGAAATAT CCGTTCAAAC TTGGAGCAGC GTTACGGCAA AGAAAACATC
2101 ACCTCCTCAA CCGTGCCGCC GTCAAACGGC AAAAATGTCA AACTGGCAGA
2151 CCAACGCCAC CCGAAGACAG GCGTACCGTT TGACGGTAAA GGGTTTCCGA
2201 ATTTTGAGAA GCACGTGAAA TATGATACGC TCGAGCACCA CCACCACCAC
2251 CACTGA

1 MATNDDDVKK AATVAIAAAY NNGQEINGFK AGEITYDIDE DGTITKDAT
51 AADVEADDFK GLGLEKVVVN LTKTVNENKQ NVDAKVKAEE SEIEKL/TKL
101 ADTDAALADT DAALDATTNA LNKLGENTTT FARETKTNIV KIDEKLEAVA
151 DTVDRKHAFA NDIADSLDET NTKADEAVKT ANEAKQTAEK TKQNVDAKVK
201 AABTAAGKAE AAAGTANTAA DKAEVAANKV TDIKADIATN KDNIAKKANS
251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDKT
301 VSDLRKETRQ GLAEQAALSG LFQPYNVGGS GGGGSDLAND SFIRQVLDRO
351 HPEPDGKYHL FGSRGELARR SGHIGLGRIQ SHQLGNLMIQ QAAIKGNIGY
401 IVRFSDHGHE VHSPPDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD
451 GYDGPQGGGY PAPKGARDY SYDIKGVQON IRLNLTDNRS TGQRLADRFH
501 NAGSMLTQGV GDGFKRATRY SPELDRSGNA ARAFNGTADI VKNIIGAAGE
551 IVGAGDAVQG ISEGSNLAVM HGLGLLSTEN KMARINDLAD MAQLKDZAAA
601 AIRDWAQNP NAAQGLEAVS NIFMAAIPK GIGAVRGKYG LGGITAEPIK
651 RSQMGALALP KGSASVDMF ADAAYAKYPS PYHSRNIRSN LBQRYKQENI
701 TSSTVPPSNG KNVKLADQRH PKTGVPPDGK GFFPNFEKVK YDTLEHHHHH

751 H*

961c-741

```

1 ATGGCCACAA ACGACGACGA TGTAAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCCCTAC AACAAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTTCATT
501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG
551 CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACCTCGATGC CAAAGTAAAA
601 GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
651 TACTGCAGCC GACAAGGCCG AGCTGTGCGC TGCAAAAGTT ACCGACATCA
701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
751 GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTTG TCAGAAATGA
801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
851 AAAAAATCCAT TGCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
901 GTGTCAGACC TGCGCAAGA AACC CGCCAA GGCCTTGACG AACAAAGCCG
951 GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTGGATCC GGAGGGGGTG
1001 GTGTCGCCGC CGACATCGGT GCGGGGCTTG CCGATGCACT AACCGCACCG
1051 CTCGACCATA AAGACAAGG TTGCAATCT TTGACGCTGG ATCAGTCCGT
1101 CAGGAAAAAC GAGAACTGA AGCTGGCGGC ACAAGGTGCG GAAAAAATCT
1151 ATGGAACCGG TGACAGCCTC AATACGGGCA AATTGAAGAA CGACAAGGTC
1201 AGCCGTTTCG ACTTTATCCG CCAAATCGAA GTGGACGGGC AGCTCATTAC
1251 CTTGGAGAGT GGAGAGTTC AAGTATACAA ACAAGCCAT TCCGCCTTAA
1301 CCGCCTTTCA GACCGAGCAA ATACAAGAT CCGAGCATTC CGGGAGATG
1351 GTTGCGAAC GCCAGTTCAG AATCGGCGAC ATAGCGGCG AACATACATC
1401 TTTTGACAAG CTCCCGAAG GCGGCAGGCG GACATATCGC GGGACGGCGT
1451 TCGGTTTCTG CGATGCCGGC GGAAGACTGA CCTACACCAT AGATTTCGCC
1501 GCCAAGCAGG GAAACGGCAA AATCGAACAT TTGAAATCGC CAGAACTCAA
1551 TGTCGACCTG GCCGCCGCCG ATATCAAGCC GGATGGAAAA CGCCATGCCG
1601 TCATCAGCGT TTCCGTCCTT TACAACCAAG CCGAGAAAGG CAGTTACTCC
1651 CTCGCTATCT TTGGCGGAAA AGCCAGGAAA GTTGCCGGCA GCGCGGAAGT
1701 GAAACCGTA AACGGCATAC GCCATATCGG CCTTGCCGCC AAGCAACTCG
1751 AGCACCACCA CCACCACCAC TGA

```

```

1 MATNDDVKK AATVALAAAY NNGQINGFK AGETIYDIDE DGTITEKDAT
51 AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDKVKAAE SEIRKLTTKL
101 ADTDALADT DAALDATNA LNKLGENTT FAKETKNTV KIDEKLEVA
151 DTVDKHABAF NDIADSLDET NTKADBAVKT ANEAKQTAEB TKQNVDAKVK
201 AAEATAGKAE AAAGTANTAA DKAEVAKV TDIKADIATN KDNIAKANS
251 ADVYTRBSQ SKFVRIDGLN ATTEKLDTRL ASAEKSIADH DTRLNGLDRT
301 VSDLRKETRQ GLAEQAALSG LFQFYNVGGG GGGVAAADIG AGLADALTAP
351 LDHKDKGLQS LTLDSVRKN EKLKLAQGA EKYNGGDSL NTGKLNDKV
401 SRFDFIRQIE VDGQLITLES GEFQVYKQSH SALTAFTQEQ IQDSEHSGKM
451 VAKRQFRIGD IAGEHTSFDK LPEGGRTYR GTAFGSDDAG GKLYTTIDFA
501 AKQNGKIEH LKSPFLNVDL AADEKPDGK REAVIGSVL YNQAENGYSYS
551 LGIFGGAQAE VAGSAEVKTV NGIRHIGLAA KQLEHHHHHH *

```

961c-983

```

1 ATGGCCACAA ACGACGACGA TGTAAAAAA GCTGCCACTG TGGCCATTGC
51 TGCTGCCCTAC AACAAATGGCC AAGAAATCAA CGGTTTCAAA GCTGGAGAGA
101 CCATCTACGA CATTGATGAA GACGGCACAA TTACCAAAAA AGACGCAACT
151 GCAGCCGATG TTGAAGCCGA CGACTTTAAA GGTCTGGGTC TGAAAAAAGT
201 CGTGACTAAC CTGACCAAAA CCGTCAATGA AAACAAACAA AACGTCGATG
251 CCAAAGTAAA AGCTGCAGAA TCTGAAATAG AAAAGTTAAC AACCAAGTTA
301 GCAGACACTG ATGCCGCTTT AGCAGATACT GATGCCGCTC TGGATGCAAC
351 CACCAACGCC TTGAATAAAT TGGGAGAAAA TATAACGACA TTTGCTGAAG
401 AGACTAAGAC AAATATCGTA AAAATTGATG AAAAATTAGA AGCCGTGGCT
451 GATACCGTCG ACAAGCATGC CGAAGCATTC AACGATATCG CCGATTTCATT
501 GGATGAAACC AACACTAAGG CAGACGAAGC CGTCAAAACC GCCAATGAAG

```

```

551 CCAAACAGAC GGCCGAAGAA ACCAAACAAA ACGTCGATGC CAAAGTAAAA
601 GCTGCAGAAA CTGCAGCAGG CAAAGCCGAA GCTGCCGCTG GCACAGCTAA
651 TACTGCAGCC GACAAGGCCG AAGCTGTGCG TGCAAAAGTT ACCGACATCA
701 AAGCTGATAT CGCTACGAAC AAAGATAATA TTGCTAAAAA AGCAAACAGT
751 GCCGACGTGT ACACCAGAGA AGAGTCTGAC AGCAAATTGT TCAGAAATTGA
801 TGGTCTGAAC GCTACTACCG AAAAATTGGA CACACGCTTG GCTTCTGCTG
851 AAAAATCCAT TCCGATCAC GATACTCGCC TGAACGGTTT GGATAAAACA
901 GTGTCAGACC TCCGCAAAGA AACCCGCCAA GGCTTGCGAG AACAAAGCCG
951 GCTCTCCGGT CTGTTCCAAC CTTACAACGT GGGTGGATCC GCGGAGGGCG
1001 GCACTTCTGC GCCCGACTTC AATGCAGGCG GTACCGGTAT CCGCAGCAAC
1051 AGCAGAGCAA CAACAGCGAA ATCAGCAGCA GTATCTTACG CCGGTATCAA
1101 GAACGAATAG TCCAAAGACA GAAGCATGCT CTGTCCCGGT CCGGATGACG
1151 TTGCGGTTAC AGCAGGGAT GCCAAATCA ATGCCCCCCC CCCGAATCTG
1201 CATACCGGAG ACTTTCAAA CCCAAATGAC GCATACAAGA ATTTGATCAA
1251 CCTCAACCT GCATTAAG CAGGCTATAC AGGACGCGGG GTAGAGGTAG
1301 GTATCGTCGA CACAGGCGAA TCCGTCCGCA GCATATCCTT TCCCAGACTG
1351 TATGGCAGAA AAGAACACGG CTATAACGAA AATTACAAA ACTATACGGC
1401 GTATATCGCG AAGGAAGCGC CTGAAGACGG AGGCGGTAAA GACATTGAAG
1451 CTTCTTTTGA CGATGAGGCC GTTATAGAGA CTGAAGCAA GCCGACGGAT
1501 ATCCGCCACG TAAAGAAAT CCGACACATC GATTTGGTCT CCCATATTAT
1551 TGGCGGGCGT TCCGTGGACG CGAGACCTGC AGGCGGTATT GCGCCCGATG
1601 CGACGCTACA CATAATGAAT ACGAATGATG AAACCAAGAA CGAAATGATG
1651 GTTGCAAGCA TCCCAATGC ATGGGTCAAG CTGGGCGAAG GTGGCGTGGC
1701 CATCGTCAAT AACAGTTTGG GAACAACATC GAGGGCAGGC ACTGCCGACC
1751 TTTTCCAAAT AGCCAATTGC GAGGAGCAGT ACCGCCAAGC GTTGCTCGAC
1801 TATTCCGGCG GTGATAAAAC AGACGAGGGT ATCCGCTGA TGCAACAGAG
1851 CGATTACGGC AACCTGTCTT ACCACATCCG TAATAAAAAC ATGCTTTTCA
1901 TCTTTTCGAC AGGCAATGAC GCACAAGCTC AGCCCAACAC ATATGCCCTA
1951 TTGCCATTTT ATGAAAAAGA CGCTCAAAA GGCATTATCA CAGTCGCAGG
2001 COTAGACCGC AGTGGAGAAA AGTTCAAACG GGAAATGTAT GGAGAACCGG
2051 GTACAGAACC GCTTGAGTAT GGCTCCAACC ATTGCGGAAT TACTGCCATG
2101 TGGTGCTGTG CGGCACCCTA TGAAGCAAGC GTCCGTTTCA CCCGTACAAA
2151 CCCGATTCAA ATTCCGGAA CATCCTTTTC CGCACCCATC GTAACCGCA
2201 CCGCGGCTCT GCTGCTGCAG AAATACCCGT GGATGAGCAA CGACAACCTG
2251 CGTACCACGT TGCTGACGAC GGCTCAGGAC ATCGGTGCAG TCGGCGTGGG
2301 CAGCAAAGTT GCGTGGGGAC TGCTGGATGC GGGTAAGGCC ATGAACGGAC
2351 CCGCGTCTCT TCCGTTCGGC GACTTTACCG CCGATACGAA AGGTACATCC
2401 GATATTCGCT ACTCCTTCGG TAACGACATT TCAGGCACGG GCGGCCTGAT
2451 CAAAAAAGGC GGCAGCCAAC TGCAACTGCA CGGCAACAAC ACCTATACGG
2501 GCAAAACCAT TATCGAAGGC GGTTCGCTGG TGTTGTACGG CAACAACAAA
2551 TCGGATATGC CGCTCGAAAC CAAAGGTGCG CTGATTTATA ACGGGGCGCG
2601 ATCCGCGCGC AGCCTGAACA GCGACGGCAT TGTCTATCTG GCAGATACCG
2651 ACCAATCCGG CGCAAACGAA ACCGTACACA TCAAAGGCAG TCTGCAGCTG
2701 GACGGCAAAG GTACGCTGTA CACACGTTTG GGCAAACCTG TGAAAGTGA
2751 CGGTACGGCG ATTATCGGCG GCAAGCTGTA CATGTCCGCA CCGGCAAGG
2801 GGGCAGGCTA TCTCAACAGT ACCGACGAC GTGTTCCTT CCTGAGTGCC
2851 GCCAAAATCG GCGAGGATTA TTCTTTCTTC ACAAACATCG AAACCGACGG
2901 CCGCCTGCTG GCTTCCCTCG ACAGCGTCGA AAAAACAGCG GGCAGTGAAG
2951 GCGACACGCT GTCCTATTAT GTCCGTGCGG GCAATGCGGC ACGGACTGCT
3001 TCGGACGCGG CACATTCCGC GCCCGCCGGT CTGAAACACG CCGTAGAACA
3051 GGGCGGCAGC AATCTGGAAG ACCTGATGGT CGAACTGGAT GCCTCCGAAT
3101 CATCCGCAAC ACCCGAGACG GTTGAAACTG CCGCAGCCGA CCGCACAGAT
3151 ATGCCGGGCA TCCGCCCTTA CCGCGCAACT TTCCGCGCAG CCGCAGCCGT
3201 ACAGCATGCG AATGCCGCCG ACGGTGTACG CATCTTCAAC AGTCTCGCCG
3251 CTACCGTCTA TGCCGACAGT ACCGCCGCC ATGCCGATAT GCAGGGACGC
3301 CGCCTGAAAG CCGTATCGGA CCGGTTGGAC CACAACGGCA CCGGTCTGCG
3351 CGTCATCGCG CAAACCCAAC AGGACGGTGG AACGTGGGAA CAGGGCGGTG
3401 TTGAAGGCAA AATGCCGCGC AGTACCCAAA CCGTCGGCAT TGCCGCGAAA
3451 ACCGGCGAAA ATACGACAGC AGCCGCCACA CTGGGCATGG GACGCAGCAC
3501 ATGGAGCGAA AACAGTGCAA ATGCAAAAAC CGACAGCATT AGTCTGTTTG
3551 CAGGCATACG GCACGATCGG GCGGATATCG GCTATCTCAA AGGCCTGTTT
3601 TCCTACGGAC GCTACAAAA CAGCATCAGC CGCAGCACC GTGCGGACGA
3651 ACATGCGGAA GCGAGCGTCA ACGGCACGCT GATGCAAGTG GCGCACTGG
3701 GCGGTGTCAA CGTTCGTTT GCCGCAACGG GAGATTTGAC GGTGGAAGGC
3751 GGTCTGCGCT ACGACCTGCT CAAACAGGAT GCATTCCCGG AAAAAGGCAG
3801 TGCTTTGGGC TGAGCGGCGA ACAGCCTCAC TGAAGGCAG CTGGTGGGAC
3851 TCGCGGTCT GAAGCTGTG CAACCTTGA GCGATAAAGC CGTCTGTTT

```

```

3901 GCAACGGCGG GCGTGGAAACG CGACCTGAAC GGACGCGACT ACACGGTAAC
3951 GGGCGGCTTT ACCGGCGCGA CTGCAGCAAC CGGCAAGACG GGGGCACGCA
4001 ATATGCCGCA CACCCGCTCTG GTTGCCGGCC TGGCGCGGGA TGTCGAATTC
4051 GGCAACGGCT GGAACGGCTT GGCACGTTAC AGCTACGCCG GTTCCAAACA
4101 GTACGGCAAC CACAGCGGAC GAGTCGGCGT AGGCTACCGG TTCCTCGAGC
4151 ACCACCACCA CCACCACTGA

```

```

1 MATNDDDVVK AATVAIAAAAY NNGQHINGFK AGETIYDIDE DGTITEKDAT
51 AADVEADDFK GLGLKKVVTN LTKTVNENKQ NVDAKVKAAB SEIEKLTTKL
101 ADTDAALADT DAALDATMA LNLGENITT FAETKTNIV KIDEKLEAVA
151 DTVDKHAEAF NDIADSLDET NTKADEAVKT ANEAKQTAE TEQNVDKVK
201 AAETAAGKAE AAAGTANTAA DKAEVAKAV TDIKADIATN KDNIAKANS
251 ADVYTREESD SKFVRIDGLN ATTEKLDTRL ASAKSIADH DTRLNGLDKT
301 VSDLRKETRQ GLAEQAALSG LFPQYNVGGG GGGGTSAPDF NAGGTGIGSN
351 SRATTAKSAA VSYAGIKNEM CKDRSMLCAG RDDVAVTRD AKINAPPPNL
401 HTGDFFPNPD AYKNLINLKP AIEAGYTRG VEVGIVDTGE SVGSISFFEL
451 YGRKEHGYNE NYKNYTAYMR KEAPEDGGGK DREASFDEA VITEAKPTD
501 IRHVKEIGHI DLVSHIIGR SVDGRPAGGI APDATLHIMN TNDFTKMM
551 VAAIRNAWVK LGERGVRIVN NSFGTTSRAG TADLFQIANS EEQYRQALLD
601 YSGGDKTDEG IRLMQSDYG NLSYHIRNKN MLFIFSTGND AQAQPMYAL
651 LPPYEDDAQK GIITVAGVDR SGEKFKREMY GEPGTEPLEY GSNHCGITAM
701 WCLSAFYEAS VKFTRTNPIQ IAGTSFSAPI VTGTAALLLQ KYPMMENDNL
751 RTTLLTTAQD IGAVGVDSKF GWGLLDAGKA MNGPASFPFG DFTADTKGTS
801 DIAYSFRNDI SGTGGLIKKG GSQQLHGMN TYTGKTIIEG GSLVLYGNK
851 DMRVETKGA LTYNGAASGG SLNSDGIVYL ADTDQSGANE TVHKGSLQL
901 DGKGTLYTRL GKLLKVDGTA IIGGKLYMSA RKGAGYLSN TGRRVFPLSA
951 AKIGQDYSFF TNIETDGLL ASLDSVEKTA GSEGDLSY VRRGNAARTA
1001 SAAHSAAPAG LKHAVEQGS NLENLMVELD ASSESTPET VETAADRTD
1051 MPGIRPYGAT FRAAAAVQHA NAADGVRIFN SLAATVYADS TAAHADMQGR
1101 RLKAVSDGLD HNGTGLRVIA QTQDDGTWE QGGVEGKMRG STQTGIAAK
1151 TGENTTAAAT LGMGRSTWSE NSANAKTDSI SLFAGIRHDA GDIGYLEGLF
1201 SYGRYKNSIS RSTGADEHAE GSVNOTLMQL GALGGVNVVF AATGDLTVEG
1251 GLRYDLLKQD AFAEKGSALG WSGNSLTGRT LVGLAGLKL QPLSDKAVLF
1301 ATAGVERDLN GRDYTVTGGF TGATAATGKT GARNMPHTRL VAGLGADVEF
1351 GNGWNLARY SYAGSKQYGN HSGRVGVGYR FLEHHHHHH*

```

961cL-ORF46.1

```

1 ATGAACACT TTCCATCCAA AGTACTGACC ACAGCCATCC TTGCCACTTT
51 CTGTAGCGG GCCTGGCAG CCACAAACGA CGACGATGTT AAAAAAGCTG
101 CCCTGTGGC CATTGCTGCT GCCTACAACA ATGGCCAAGA AATCAACGGT
151 TTCAAAGCTG GAGAGACCAT CTACGACATT GATGAAGACG GCACAATTAC
201 CAAAAAGAC GCAACTGCAG CCGATGTTGA AGCCGACGAC TTTAAAGGTC
251 TGGGTCTGAA AAAAGTCTG ACTAACCTGA CCAAAACCGT CAATGAAAAC
301 AAACAAACG TCGATGCCAA AGTAAAGCT GCAGAACTCG AAATAGAAAA
351 GTTAACAAAC AAGTTAGCAG ACACTGATGC CGCTTTAGCA GATACTGATG
401 CCGCTCTGGA TGCAACCACC AACGCCCTGA ATAAATTGGG AGAAATATA
451 ACGACATTTG CTGAAGAGAC TAAGACAAAT ATCGTAAAAA TTGATGAAAA
501 ATTAGAAGCC GTGGCTGATA CCGTCGACAA GCATGCCGAA GCATTCAACG
551 ATATCGCCGA TTCATTGGAT GAAACCAACA CTAAGGCAGA CGAAGCCGTC
601 AAAACCGCCA ATGAAGCCAA ACAGACGCCG GAAGAAACCA AACAAAACGT
651 CGATGCCAAA GTAAAGCTG CAGAACTGC AGCAGGCAAA GCCGAAGCTG
701 CCGCTGGCAC AGCTAATACT GCAGCCGACA AGGCCGAAGC TGTGCTGCA
751 AAAGTTACCG ACATCAAAGC TGATATCGCT ACGAACAAAG ATAATATTGC
801 TAAAAAGCA AACAGTGCCG ACGTGTACAC CAGAGAAGAG TCTGACAGCA
851 AATTGTCTAG AATTGATGGT CTGAACGCTA CTACCGAAAA ATTGGACACA
901 CGCTTGGCTT CTGCTGAAAA ATCCATTGCC GATCACGATA CTCGCCGAA
951 CGGTTTGGAT AAAACAGTGT CAGACCTGCG CAAAGAAACC CGCCAAGGCC
1001 TTGCAGAAAC AGCCCGCTC TCCGGTCTGT TCCAACCTTA CAACGTGGGT
1051 GGATCCGGAG GAGGAGGATC AGATTGGCA AACGATTCTT TTATCCGGCA
1101 GGTCTCTGAC CGTCAGCATT TCGAACCAGA CGGGAAATAC CACCTATTCC
1151 GCAGCAGGGG GGAACCTGCC GAGCGCAGCG GCCATATCGG ATTGGGAAAA
1201 ATACAAAGCC ATCAGTTGGG CAACCTGATG ATTCAACAGG CGGCCATTAA
1251 AGGAAATATC GGCTACATTG TCCGCTTTTC CGATCACGGG CACGAAGTCC
1301 ATTCCCCCTT CGACAACCAT GCCTCACATT CCGATTCTGA TGAAGCCGGT
1351 AGTCCCGTTG ACGGATTTAG CTTTACCOC ATCCATTGGG ACGGATACGA
1401 ACACCATCCC GCCGACGGCT ATGACGGGCC ACAGGGCGGC GGCTATCCCG

```

```

1451 CTCCCAAAGG CGCGAGGGAT ATATACAGCT ACGACATAAA AGGCGTTGCC
1501 CAAAATATCC GCCTCAACCT GACCGACAAC CGCAGCACCG GACAAACGGCT
1551 TGCCGACCGT TPCACAAATG CCGGTAGTAT GCTGACGCAA GGAGTAGGCG
1601 ACGGATTCAA ACGCGCCACC CGATACAGCC CCGAGCTGGA CAGATCGGGC
1651 AATGCCGCCG AAGCCTTCAA CGGCACTGCA GATATCGTTA AAAACATCAT
1701 CGGCGCGGCA GGAGAAATTG TCGGCGCAGG CGATGCCGTG CAGGGCATAA
1751 GCGAAGGCTC AAACATTGCT GTCATGCACG GCTTGGGTCT GCTTTCACC
1801 GAAAACAAGA TGGCGCGCAT CAACGATTG GCAGATATGG CGCAACTCAA
1851 AGACTATGCC GCAGCAGCCA TCCGCGATTG GGCAGTCCAA AACCCCAATG
1901 CCGCACAAGG CATAGAAGCC GTCAGCAATA TCTTTATGGC AGCCATCCCC
1951 ATCAAAGGGA TTGGAGCTGT TCGGGGAAAA TACGGCTTGG GCGGCATCAC
2001 GGCACATCCT ATCAAGCGGT CGCAGATGGG CGCGATCGCA TTGCCGAAAG
2051 GGAAATCCCG CGTCAGCGAC AATTTTGCCG ATGCGGCATA CGCCAAATAC
2101 CCGTCCCTT ACCATTCCCG AAATATCCGT TCAAACCTGG AGCAGCGTTA
2151 CGGCAAAGAA AACATCACCT CCTCAACCGT GCCGCCGTCA AACGGCAAAA
2201 ATGTCAAACG GGCAGACCAA CGCCACCCGA AGACAGGCGT ACCGTTTGAC
2251 GGTAAAGGGT TTCCGAATTT TGAGAAGCAC GTGAAATATG ATACGTAAC
2301 CGAG

```

```

1 MKHFPSKVL TAILATFCSG ALAATNDDDV KKAATVAIAA AYNNGQHING
51 FKAGETIYDI DEDGTITKED ATAADV EADD FKGLGLKVV TNLTKTVNEN
101 KQNVDAKVA AESIEKLT KLADTDAALA DTDALDATT NALNKLGENI
151 TTFABETKTN IVKIDKLEA VADTVDKHAE AFMDIADSLD BTNTKADAEV
201 KTANEAKQTA ESTKQNVDAK VKAABTAAGK AEAAAGTANT AADKABAVAA
251 KVTDIKADIA TNKDNIAKKA NSADVITREE SDSKFVRIDG LNATTEKLDI
301 RLASAESKIA DHDTRLNGLD KTVSDLRKET RQGLABQAL SGLFPQYINV
351 GSGGGGSDLA NDSFIRQVLD RQHFEPDGKY HLFPSRGELA ERSGHIGLKK
401 IQSHQLGNLM IQQAALIGNI GYIVRFSDHG HBVHSPFDNH ASHSDSDEAG
451 SPVDGFSLYR IHMDGYEHHP ADGYDGFQGG GYPAPKARD IYSYDIKQVA
501 QNIRLNLTDN RSTGQRLADR FHNAGSMLTQ GVGDGFKRAT RYSPFLDRSG
551 NAABAFNGTA DIVKHIIGAA GEIVGAGDAV QGISSEGSNIA VMHGLQLLST
601 ENEMARINDL ADMAQLKDYA AAAIRDWAVQ NPMAAQGLEA VSNIFMAAIF
651 IKIGIAVRGK YLOGGITAHF IKRSQMGATA LPRGKSASVD NFADAAYAKY
701 PSPYHSRIR SNLEQRYGKE NITSSTVPPS NGKNVKLADQ RHPKRGVPPD
751 GKGFPPFEKH VKYDT*

```

961cL-741

```

1 ATGAAACACT TTCCATCCAA AGTACTGACC ACAGCCATCC TTGCCACTTT
51 CTGTAGCGGC GCACTGGCAG CCACAAACGA CGACGATGTT AAAAAAGCTG
101 CCACTGTGGC CATGTCTGCT GCCTACAACA ATGCCCAAGA AATCAACGGT
151 TTCAAAGCTG GAGAGACCAT CTACGACATT GATGAAGACG GCACAATTAC
201 CAAAAAAGAC GCAACTGCAG CCGATGTTGA AGCCGACGAC TTAAAGGCTC
251 TGGGTCTGAA AAAAGTCTGT ACTAACCTGA CCAAAACCGT CAATGAAAAC
301 AAACAAAACG TCGATGCCAA AGTAAAAGCT GCAGAATCTG AATAGAAAA
351 GTTAACAACC AAGTTAGCAG ACACTGATGC CGCTTTAGCA GATACTGATG
401 CCGCTCTGGA TGCAACCACC AACGCCTTGA ATAAATTGGG AGAAAAATA
451 ACGACATTTG CTGAAGAGAC TAAGACAAAT ATCGTAAAAA TTGATGAAAA
501 ATTAGAAGCC GTGGCTGATA CCGTCGACAA GCATGCCGAA GCATTCAACG
551 ATATCGCCGA TTCATTGGAT GAAACCAACA CTAAGGCAGA CGAAGCCGTC
601 AAAACCGCCA ATGAAGCCAA ACAGACGGCC GAAGAAACCA AACAAAACGT
651 CGATGCCCAA GTAAAAGCTG CAGAAACTGC AGCAGGCAAA GCCGAAGCTG
701 CCGCTGGCAC AGCTAATACT GCAGCCGACA AGGCCGAAGC TGTCTGTGCA
751 AAAGTTACCG ACATCAAAGC TATATCGCT ACGAACAAAG ATAATATTGC
801 TAAAAAAGCA AACAGTGCCG ACGGTACAC CAGAGAAGAG TCTGACAGCA
851 AATTGTCTAG AATTGATGGT CTGAACGCTA CTACCGAAAA ATTGGACACA
901 CGCTTGGCTT CTGCTGAAAA ATCCATTGCC GATCAGGATA CTCGCCTGAA
951 CGGTTTGGAT AAAACAGTGT CAGACCTGCG CAAAGAAACC CGCCAAGGCC
1001 TTGCAGAAC AAGCCGCGCTC TCCGGTCTGT TCCAACCTTA CAACGTGGGT
1051 GGATCCGGAG GGGGTGGTGT CGCCGCCGAC ATCGGTGCGG GGCTTGCCGA
1101 TGCACTAACC GCACCGCTCG ACCATAAAGA CAAAGGTTTG CAGTCTTTGA
1151 CGCTGGATCA GTCCGTCAGG AAAAACGAGA AACTGAAGCT GCGGCGACAA
1201 GGTGCGGAAA AAACCTATGG AACGGGTGAC AGCCTCAATA CGGGCAAAAT
1251 GAAGAACGAC AAGGTCAGCC GTTTCGACTT TATCCGCCAA ATCGAAGTGG
1301 ACGGGCAGCT CATTACCTTG GAGAGTGGAG AOTTCCAAGT ATACAAACAA
1351 AGCCATTCCG CCTTAACCGC CTTTCAGACC GAGCAAATAC AAGATTCCGA
1401 GCATTCCGGG AAGATGGTTG CGAAACGCCA GTTCAGAATC GGCACATAG

```

```

1451 CCGGCGAACA TACATCTTTT GACAAGCTTC CCGAAGGCGG CAGGGCGACA
1501 TATCGCGGGA CGGCGTTCGG TTCAGACGAT GCCGGCGGAA AACTGACCTA
1551 CACCATAGAT TTCGCCGCGA AGCAGGGAAA CGGCAAAATC GAACATTTGA
1601 AATCGCCAGA ACTCAATGTC GACCTGGCCG CCGCCGATAT CAAGCCGGAT
1651 GGAAACGCC ATGCCGTCAT CAGCGGTTCG GTCCTTTACA ACCAAGCCGA
1701 GAAAGGCAGT TACTCCCTCG GTATCTTTGG CGGAAAAGCC CAGGAAGTTG
1751 CCGGCAGCGC GGAAGTGAAA ACCGTAAACG GCATACGCCA TATCGGCCTT
1801 CCCGCCAAGC AACTCGAGCA CCACCACCAC CACCACTGA

1 MKHFPSKVL TAILATFCSG ALAATNDDUV KKAATVAIAA AYNINGQRIING
51 FKAGETIYDI DEDGTITKKD ATAADVEADD FKGLGLKKVV TNLTKTVNEN
101 KQNVDAKVKA AESEIEKLTT KLADTDAALA DTDALDATT NALNKLGENI
151 TTFABETKTN IVKIDKLEA VADTVDKHAE AFNDIADSLD ETNTKADKAV
201 KTANEAKQTA EETKQNVDAK VKAABTAAGK AEAAAAGTANT AADKABAVAA
251 KVTDIKADIA TNKDNIAKKA NSADVYTREE SDSKFVRIDG LNATTEKLDI
301 RLASAEKSIA DHDTRLNGLD KTVSDLRKET RQGLAEQAAL SGLFQFYNVG
351 GSGGGVVAAD IGAGLADALT APLDHKDKGL QSLTLDQSVR KNEKLELAAQ
401 GAETTYGNGD SLMTGKLKND KVSRLFDFIRQ IEVDGQLITL ESGEFQVYKQ
451 SHSALTAFQT BQIQDSHESG KMWAKRQFRI GDIAGEHTSF DKLPBGGRAT
501 YRGTAFGSDO AGGKLTYYTID FAAKQNGSKI EHLKSPFLNV DLAAADIKPD
551 GKRHAVISGS VLYNQAEKGS YSLGIFGKA QEVAGSAEVK TVNGIRHIGL
601 AAKQLEHHHH HH*

```

961cL-983

```

1 ATGAAACACT TTCCATCCAA AGTACTGACC ACAGCCATCC TTGCCACTTT
51 CTGTAGCGGC GCACTGGCAG CCACAAACGA CGACGATGTT AAAAAAGCTG
101 CCATGTGGC CATGTCTGCT GCCTACAACA ATGGCCAAGA AATCAACGGT
151 TTCAAAGCTG GAGAGACCAT CTACGACATT GATGAAGACG GCACAATTAC
201 CAAAAAGAC GCAACTGCAG CCGATGTTGA AGCCGACGAC TTAAAGGTC
251 TGGGTCTGAA AAAAGTCGTG ACTAACCCTGA CCAAAACCGT CAATGAAAAC
301 AAACAAACG TCGATGCCAA AGTAAAGCT GCAGAATCTG AAATAGAAAA
351 GTTAACAACC AAGTTAGCAG ACACTGATGC CGCTTTAGCA GATACTGATG
401 CCGCTCTGGA TGCAACCACC AACGCCCTGA ATAAATTGGG AGAAAATATA
451 ACGACATTG CTGAAGAGAC TAAGACAAAT ATCGTAAAAA TTGATGAAAA
501 ATTAGAAGCC GTGCTGATA CCGTCGACAA GCATGCCGAA GCATTCAACG
551 ATATCGCCGA TTCATTGGAT GAAACCAACA CTAAGGCAGA CGAAGCCGTC
601 AAAACCGCCA ATGAAGCCAA ACAGACGGCC GAAGAAACCA AACAAAACGT
651 CGATGCCAAA GTAAAGCTG CAGAACTGC AGCAGGCAAA GCCGAAGCTG
701 CCGCTGGCAC AGCTAATACT GCAGCCGACA AGGCCGAAGC TGTCGCTGCA
751 AAAGTTACCG ACATCAAAGC TGATATCGCT ACGAACAAAG ATAATATTGC
801 TAAAAAGCA AACAGTGCCG ACGTGTACAC CAGAGAAGAG TCTGACAGCA
851 AATTTGTGAG AATTGATGGT CTGAACGCTA CTACCGAAAA ATTGGACACA
901 CGCTTGCTT CTGCTGAAAA ATCCATTGCC GATCAGGATA CTGCGCTGAA
951 CGGTTTGGAT AAAACAGTGT CAGACCTGCG CAAAGAAACC CGCCAAGGCC
1001 TTGCAGAAC AGCCGCGCTC TCCGGTCTGT TCCACCTTA CAACGTGGGT
1051 GGATCCGGCG GAGGCGGCAC TTCTGCGCCC GACTTCAATG CAGGCGGTAC
1101 CGGTATCGGC AGCAACAGCA GAGCAACAAC AGCGAAATCA GCAGCAGTAT
1151 CTTACGCCGG TATCAAGAAC GAAATGTGCA AAGACAGAAG CATGCTCTGT
1201 GCCGGTGGG ATGACGTTGC GGTACAGAC AGGGATGCCA AAATCAATGC
1251 CCCCCCCC AATCTGCATA CCGGAGACTT TCCAAACCCA AATGACGCAT
1301 ACAAGAAATT GATCAACCTC AAACCTGCAA TTGAAGCAGG CTATACAGGA
1351 CCGGGGTTAG AGGTAGGTAT CGTCGACACA GCGGAATCCG TCGGCAGCAT
1401 ATCCTTTCCC GAACGTATAT GCAGAAAAGA ACACGGCTAT AACGAAAATT
1451 AAAAAAATA TACGGCGTAT ATGCGGAAGG AAGCGCCTGA AGACGGAGGC
1501 GGTAAAGACA TTGAAGCTTC TTTCGACGAT GAGGCCOTTA TAGAGACTGA
1551 AGCAAAGCCG ACGGATATCC GCCACGTAAA AGAAATCGGA CACATCGATT
1601 TGGTCTCCCA TATTATGGC GGGCGTTCCG TGGACGGCAG ACCTGCAGGC
1651 GGTATTGGCG CCGATGCGAC GCTACACATA ATGAATACGA ATGATGAAC
1701 CAAGAACGAA ATGATGGTTG CAGCCATCCG CAATGCATGG GTCAAGCTGG
1751 GCGAACGTGG CGTGGCGATC GTCAA*AA CA GTTTTGGAAC AACATCGAGG
1801 GCAGGCACTG CCGACCTTTT CCAATAGCC AATTCCGAGG AGCAGTACCG
1851 CCAAGCGTTG CTCGACTATT CCGCGGTGTA TAAACAGAC GAGGGTATCC
1901 GCGTGTATGCA ACAGAGCGAT TACGGCAACC TGTCTTACCA CATCCGTAAT
1951 AAAAAATGC TTTTCATCTT TTCGACAGGC AATGACGCAC AAGCTCAGCC
2001 CAACACATAT GCCCTATTGC CATTTTATGA AAAAGACGCT CAAAAAGGCA
2051 TTATCACAGT CGCAGGCGTA GACCGCAGTG GAGAAAAGTT CAAACGGGAA

```

```

2101 ATGTATGGAG AACCGGTAC AGAACCGCTT GAGTATGGCT CCAACCATTG
2151 CGGAATTACT GCCATGTGGT GCCTGTCGGC ACCCTATGAA GCAAGCGTCC
2201 GTTTCACCCG TACAAACCCG ATTCAAATTG CCGGAACATC CTTTTCGCGA
2251 CCCATCGTAA CCGGCACGGC GGCTCTGCTG CTGCAGAAAT ACCCGTGGAT
2301 GAGCAACGAC AACCTGCGTA CCACGTTGCT GACGACGGCT CAGGACATCG
2351 GTGCAGTCGG COTGGACAGC AAGTTCGGCT GGGGACTGCT GGATCGGGT
2401 AAGGCCATGA ACGGACCCGC GTCTTTTCCG TTCGGCGACT TTACCGCCGA
2451 TACGAAAGGT ACATCCGATA TTGCCCTACTC CTTCCGTAC GACATTTTCAG
2501 GCACGGGCGG CCTGATCAAA AAAGGCGGCA GCCAACTGCA ACTGCACGGC
2551 AACAAACACT ATACGGGCAA AACCATTATC GAAGGCGGTT CGCTGOTGTT
2601 GTACGGCAAC AACAAATCGG ATATGCGCGT CGAAACCAA GGTGCGCTGA
2651 TTTATAACGG GCGGCATCC GCGGCAGCC TGAACAGCGA CGGCATTGTC
2701 TATCTGCGAG ATACCGACCA ATCCGCGCGA AACGAAACCG TACACATCAA
2751 AGGCAGTCTG CAGCTGGACG GCAAAGGTAC GCTGTACACA CGTTTGGGCA
2801 AACTGCTGAA AGTGGACGOT ACGGCGATTA TCGGCGGCAA GCTGTACATG
2851 TCGGCACCGG GCAAGGGGGC AGGCTATCTC AACAGTACCG GACGACGTGT
2901 TCCCTTCCTG AGTGGCGCCA AAATCGGGCA GGATTATTCT TTCTTCACAA
2951 ACATCGAAAC CGACGGCGGC CTGCTGCTT CCGTCGACAG CGTCGAAAAA
3001 ACAGCGGGCA GTGAAGGCGA CACGCTGTCC TATTATGTCC GTCGCGGCAA
3051 TCGGCGACGG ACTGCTTCGG CAGCGGCACA TTCGCGGCC GCGGCTCTGA
3101 AACACGCGT AGAACAGGGC GGCAGCAATC TGGAAAACCT GATGCTCGAA
3151 CTGGATGCCCT CCGAATCCTC CGCAACACCC GAGACGGTTG AAAGTGGGCG
3201 AGCCGACCGC ACAGATATGC CCGGCATCCG CCCCTACGGC GCAACTTTCC
3251 GCGCAGCGGC AGCCGTACAG CATGCGAATG CCGCGACCG TGTACGCATC
3301 TTCAACAGTC TCGCGCTAC CGTCTATGCC GACAGTACCG CCGCCCATGC
3351 CGATATGCA GACGCGCGCC TGAAAGCGT ATCGGACGGG TTGGACCACA
3401 ACGGCACGGG TCTGCGCGTC ATCGCGCAA CCCAACAGGA CGGTGGAACG
3451 TGGGAACAGG GCGGTGTGA AGGCAAAATG CGCGCGAGTA CCCAAACCGT
3501 CCGCATTGCC GCGAAAACCG GCGAAAATAC GACAGCAGCC GCCACACTGG
3551 GCATGGGACG CAGCACATGG AGCGAAAACA GTGCAAAATG AAAAAACGAC
3601 AGCATTAGTC TGTTTGACAG CATAACGAC GATGCGGGCG ATATCGGCTA
3651 TCTCAAAGGC CTGTTCTCCT ACGGACGCTA CAAAAACAGC ATCAGCCGCA
3701 GCACCGGTGC GACGAAACAT CCGGAAGGCA GCGTCAACGG CACGCTGATG
3751 CAGCTGGGCG CACTGGGCGG TGTCAACGTT CCGTTTGCCG CAACGGGAGA
3801 TTTGACGGTC GAAGGCGGTC TCGGCTACGA CCTGCTCAA CAGGATGCAT
3851 TCGCCGAAAA AGGCAGTGT TGGGCTGGA GCGGCAACAG CCTCACTGAA
3901 GGCACGCTGG TCGGACTCGC GGGTCTGAAG CTGTGCAAC CCTTGAGCGA
3951 TAAAGCCGTC CTGTTTGCAA CCGCGGGCGT GGAACGCGAC CTGAACGGAC
4001 GCGACTACAC GGTAAACGGG GCGTTTACCG GCGCGACTGC AGCAACCGGC
4051 AAGACGGGGG CACGCAATAT GCGCACACC CGTCTGGTTG CCGGCCTGGG
4101 CCGGATGTC GAATTCGGCA ACGGCTGGAA CCGCTTGGCA COTTACAGCT
4151 ACGCCGGTTC CAAACAGTAC GGCAACCACA GCGGACGAGT CGCGTAGGCG
4201 TACCGGTTCT GACTCGAG

```

```

1 MKHFPKSVLT TAILATFCSG ALAATNDDDV KKAATVAIAA AYNGQEIING
51 FKAGETIYDI DEDGTTTKD ATAADVADD FKGLGLKKVV TMLTKTVNEN
101 KQNVDAKVA ABSEIEKLT KLADTDAALA DTDAALDATT NALNKLGENI
151 TTFABETKTN IVKIDKLEA VADTVDKHAE AFNDIADSLD ETWTKADEAV
201 KTANEAKQTA BBTQNVDAK VKAETAAGK AEAAGTANT AADKAEAVAA
251 KVTDIKADIA TMDNTAKKA NSADVYTREE SDSKFVRIDG LNAATTEKLD
301 RLASAESKIA DHDTRLNGLD KTVSLRKET RQGLAEQAAL SGLFPQYNVG
351 GSGGGGTSAP DFNAGGTGIG SNSRATTAKS AAVSYAGIKN BMCNDRSMLC
401 AGRDDVAVTD RDAKINAPP NLHTGDFPNP NDAYKNLNL KPAIEAGYTG
451 RGVEVGIVDT GESVGSISFP ELYGRKEHGY NENYKNYTAY MRKEAPEDGG
501 GKDIASFD D EAVIETAKP TDIRHVKEIG HIDLVSHIIG GRSVDGRPAG
551 GIAPDATLHI MNTNDETKNE MMVAAIRNAN VKLGERGVRI VNNSFGTTSR
601 AGTADLFQIA NSEBQYRQAL LDYSGGDKTD EGIRLMQSD YGNLSYHIRN
651 KNMLFIFSTG NDAQAQNTY ALLPFYEKDA QKGIIITVAGV DRSGEKFKE
701 MYGEPGTEPL EYGSNHCGIT AMWCLSAPE ASVRFTRTNP IQIAGTSFSA
751 PIVTGTALL LQYFPMNSND NLRITLLTTA QDIGAVGVDS KFGWGLLDAG
801 KAMNGPASFP FGDTADTKG TSDIAYSFRN DISGTGLIK KGGSQLQLHG
851 NNTYTGTII EGGSLVLYGN NKSDMRVETK GALIYNGAAS GGSINSDGIV
901 YLADTDQSGA NETVHIKSL QLDGKOTLYT RLGLKLVKG TAIIGGKLYM
951 SARGKGAGYL NSTGERVPFL SAAKIGQDYS FFMNIETDQG LLASLDSVEK
1001 TAGSEGDLS YYVRGNAAR TASAAHSAP AGLKHAVERG GSNLENLMVE
1051 LDASSESATP ETVETAADR TDMFGIRPYG ATFRAAAAPV HANAADGVRI
1101 FNSLAATVYA DSTAAHADMQ GRRLKAVSDG LDHNGTGLRV LAQTQQDGGT

```

```

1151 WEQGGVEGIM RGSTQTVGIA AKTGENTTAA ATLMGRSTW SENSANAKTD
1201 SISLFAGIRH DAGDIGYLKG LPSYGRYKNS ISRSTGADEH AEGSVNOTLM
1251 QLQALGGVNV PFAATGDLTV EGGLFYDLLK QDAFAEKGS A LGWSGNSLTE
1301 GTLVGLAGLK LSQPLSDKAV LFATAGVERD LNGRDYTVTG GFTGATAATG
1351 KTGARNMPHT RLVAGLGADV EFGNGWNGLA RYSYAGSKQY GNHSGRVGVG
1401 YRF*

```

可以理解本发明仅以例子的方式进行描述，在本发明的范围和精神内还可进行改变。例如，设想可以使用其它菌株的蛋白质[如，参见 WO 00/66741，ORF4、ORF40、ORF46、225、235、287、519、726、919 和 953 的多聚体序列]。

实验详述

FPLC 蛋白质纯化

下表总结了所用的 FPLC 蛋白质纯化：

蛋白质	PI	柱	缓冲液	pH	方法
121.1 未标记的	6. 23	单 Q	Tris	8. 0	A
128.1 未标记的	5. 04	单 Q	Bis-Tris 丙烷	6. 5	A
406.1L	7. 75	单 Q	二乙醇胺	9. 0	B
576.1L	5. 63	单 Q	Tris	7. 5	B
593 未标记的	8. 79	单 S	Hepes	7. 4	A
726 未标记的	4. 95	Hi-trap S	Bis-Tris	6. 0	A
919 未标记的	10. 5(-前导区)	单 S	Bicine	8. 5	C
919Lorf4	10. 4(-前导区)	单 S	Tris	8. 0	B
920L	6. 92(-前导区)	单 Q	二乙醇胺	8. 5	A
953L	7. 56(-前导区)	单 S	MES	6. 6	D
982 未标记的	4. 73	单 Q	Bis-Tris 丙烷	6. 5	A
919-287	6. 58	Hi-trap Q	Tris	8. 0	A
953-287	4. 92	单 Q	Bis-Tris 丙烷	6. 2	A

10 缓冲液包括 20-120mM NaCl、5.0mg/ml CHAPS 和 10% v/v 甘油。将透析液以 13000g 离心 20 分钟，加样到单 Q 或单 S FPLC 离子交换树脂上。按照感兴趣蛋白质的 pI 和 FPLC 方法手册[pharmacia: FPLC 离子交换和层析聚焦；理论和方法。Pharmacia Publication]选择缓冲液和离子交换树脂。用分步 NaCl 梯

度法脱蛋白质。用 SDS-PAGE 分析纯化，并用 Bradford 方法测定蛋白质浓度。

‘方法’栏中的字母为以下含义：

- FPLC-A: 破碎细胞后从大肠杆菌可溶性组分中纯化克隆 121.1、128.1、593、726、982、周质蛋白质 920L 和杂交蛋白 919-287、953-287。于 37℃，在 20ml
- 5 LB/Amp(100μg/ml)液体培养基中，使携带感兴趣质粒的单集落生长过夜。用 1.0L 新鲜培养基以 1:30 稀释细菌，并在 30℃或 37℃生长直到 OD₅₅₀ 到达 0.6-0.8。以终浓度 1.0mM 的 IPTG 诱导重组蛋白质的表达。培育 3 小时后，于 4℃以 8000 离心 15 分钟收获细菌。必要时将细胞存储于-20℃。随后的所有步骤都是在冰上或 4℃进行的。对胞质蛋白质(121.1、128.1、593、726 和 982)和
- 10 周质蛋白质 920L 而言，将细菌重悬浮于 25ml 含完整蛋白酶抑制剂 (Boehringer-Mannheim)的 PBS 中。用 Branson Sonifier 450 通过超声处理裂解细胞。以 8000g 离心破碎的细胞 30 分钟，以沉淀未破碎的细胞和包含体，并加入 3.9M (NH₄)₂SO₄ 使上清液达到 35% v/v 饱和。以 8000g 沉积沉淀物 30 分钟。添加 3.9M (NH₄)₂SO₄ 使上清液达到 70% v/v 饱和，如上所述收集沉淀物。用
- 15 SDS-PAGE 鉴定含感兴趣蛋白质的沉淀物，并用适合的离子交换缓冲液(见以下)透析 6 小时或过夜。按 Evans 等[Infect.Immun.(1974)10:1010-1017]的方法制备表达 953L 大肠杆菌的周质组分，并用适合的离子交换缓冲液透析。根据感兴趣蛋白的 pI 和 FPLC 方法手册[Pharmacia]的推荐选择缓冲液和离子交换树脂。缓冲液包括 20mM NaCl 和 10%(v/v)甘油。以 13000g 离心透析液 20 分钟，并
- 20 加到单 Q 或单 S FPLC 离子交换树脂上。根据感兴趣蛋白的 pI 和 FPLC 方法手册[Pharmacia]的推荐选择缓冲液和离子交换树脂。用分步或连续 NaCl 梯度从离子交换树脂洗脱蛋白质。用 SDS-PAGE 分析纯化，并用 Bradford 方法测定蛋白质浓度。通过 NH₂-末端测序(见以下)验证周质蛋白质的前导肽的切割。

- FPLC-B: 从大肠杆菌的膜组分纯化这些蛋白质。37℃，在 20ml LB/Amp
- 25 (100μg/ml)液液培养基中使携带感兴趣质粒的单集落生长过夜。用 1.0L 新鲜培养基以 1:30 稀释细菌。克隆 406.1L 和 919LOrf4 在 30℃生长，Orf25L 和 576.1L 在 37℃生长，直到 OD₅₅₀ 达到 0.6-0.8。对 919LOrf4 而言，生长必须在 30℃，因为重组蛋白质在 37℃的表达导致细胞的裂解。以终浓度为 1.0mM 的 IPTG 诱导重组蛋白质的表达。培育 3 小时后，于 4℃以 8000g 离心 15 分钟收获细菌。

必要时，将细胞存储于-20℃。随后的所有步骤都是在冰上或4℃进行的。将细菌重悬浮于25ml含完整蛋白酶抑制剂(Boehringer-Mannheim)的PBS中，并通过French Press以渗透性冲击2-3代裂解细菌。以5000g离心15分钟除去未破碎的细胞，以100000g(Beckman Ti50, 38000rpm)离心45分钟沉淀膜。用Dounce匀浆器将膜沉淀物重悬浮于7.5ml 20mM Tris-HCl(pH 8.0)、1.0M NaCl和完整的蛋白酶抑制剂中。混合悬浮液2-4小时，以100000g离心45分钟，并将沉淀物重悬浮于7.5ml 20mM Tris-HCl(pH 8.0)、1.0M NaCl、5.0mg/ml CHAPS、10%(v/v)甘油和完整的蛋白酶抑制剂中。将此溶液混合过夜，以100000g离心45分钟，用适当选择的缓冲液透析上清液6小时。对Orf25.L而言，发现在CHAPS提取后得到的沉淀物包含重组蛋白质。无需进一步纯化，可将该组分用于免疫接种小鼠。

FPLC-C: 与FPLC-A相同，但从用多粘菌素B渗透大肠杆菌后得到的可溶性组分进行纯化，而不是在细胞破碎后进行的。

FPLC-D: 于37℃，使携带感兴趣质粒的单集落在20ml LB/Amp(100μg/ml)液态培养基中生长过夜。用1.0L新鲜培养基以1:30稀释细菌，并在30℃生长直到OD₅₅₀到达0.6-0.8。以终浓度1.0mM的IPTG诱导重组蛋白质的表达。培育3小时后，于4℃以8000g离心15分钟收获细菌。必要时将细胞存储于-20℃。随后的所有步骤都是在冰上或4℃进行的。将细胞重悬浮于20mM Bicine(pH 8.0)、20mM NaCl、10%(v/v)甘油和完整蛋白酶抑制剂(Boehringer-Mannheim)中，并用Branson Sonifier 450破碎细胞。以8000g离心超声处理液30分钟，以沉积未破碎的细胞和包含体。通过添加3.9M (NH₄)₂SO₄，从饱和度为35%v/v到70%v/v之间的溶液中沉淀重组蛋白。以8000g沉积沉淀物30分钟，将其重悬浮于20mM Bicine(pH 8.5)、20mM NaCl、10%(v/v)甘油，并用该缓冲液透析6小时或过夜。以13000g离心透析液20分钟，并加到FPLC树脂上。用分步NaCl梯度从该柱洗脱下蛋白质。用SDS-PAGE分析纯化，并用Bradford方法测定蛋白质浓度。

克隆策略和寡核苷酸设计

用以脑膜炎奈瑟球菌B MC58的基因组序列为基础设计的寡核苷酸，通过

PCR 扩增编码感兴趣抗原的基因。除非特别指出，通常将菌株 2996 的基因组 DNA 用作 PCR 反应的模板，将扩增的片段克隆入表达载体 pET21b+(Novagen)，从而以 C-末端 His 标记的产物形式表达该蛋白质，或将其克隆入 pET-24b+(Novagen)以‘未标记的’形式(如 Δ G287K)表达该蛋白质。

- 5 当蛋白质不用融合配体和用其自身前导肽(如果存在时)表达时，进行开放读框(ATG 到终止密码子)的扩增。

当蛋白质以‘未标记的’的形式表达时，通过从预计的前导序列设计 5'-端扩增引物下游除去前导肽。

- 10 用于 PCR 的引物的解链温度取决于整个引物中杂交核苷酸的数量和类型，并用以下公式确定：

$$T_{m1} = 4(G+C) + 2(A+T) \quad (\text{除去尾部的})$$

$$T_{m2} = 64.9 + 0.41(\%GC) - 600/N \quad (\text{完整的引物})$$

对整个寡聚物而言，所选寡核苷酸的解链温度通常为 65-70℃，仅针对杂交区域，解链温度为 50-60℃。

- 15 用 Perkin Elmer 394 DNA/RNA 合成仪合成寡核苷酸，将其从柱上洗脱到 2.0ml NH_4OH 中，在 56℃培育 5 小时以去保护。加入 0.3M 乙酸钠和 2 体积的乙醇沉淀寡聚物。将样品离心，将沉淀物重悬浮于水中。

		序列	限制酶切位点
Orf1L	Fwd	CGCGGATCCGCTAGC-AAAACAACCGACAAACGG	NheI
	Rev	CCCGCTCGAG-TTACCAGCGGTAGCCTA	XhoI
Orf1	Fwd	CTAGCTAGC-GGACACACTTATTTCCGCATC	NheI
	Rev	CCCGCTCGAG-TTACCAGCGGTAGCCTAATTG	XhoI
Orf1LOmpA	Fwd		NdeI-(NheI)
	Rev	CCCGCTCGAG-	XhoI
Orf4L	Fwd	CGCGGATCCCATATG-AAAACCTTCTTCAAAACC	NdeI
	Rev	CCCGCTCGAG-TTATTTGGCTGCGCCTTC	XhoI
Orf7-1L	Fwd	GCGGCATTAAT-ATGTTGAGAAAATTGTTGAAATGG	AseI
	Rev	GCGGCCCTCGAG-TTATTTTTTCAAAATATATTTGC	XhoI
Orf9-1L	Fwd	GCGGCATATG-TTACCTAACCGTTTCAAAATGT	NdeI
	Rev	GCGGCCCTCGAG-TTATTTCCGAGGTTTTCCGGG	XhoI
Orf23L	Fwd	CGCGGATCCCATATG-ACACGCTTCAAATATTC	NdeI
	Rev	CCCGCTCGAG-TTATTTAAACCGATAGGTAAA	XhoI
Orf25-1 His	Fwd	CGCGGATCCCATATG-GGCAGGGAAGAACCGC	NdeI
	Rev	GCCCAAGCTT-ATCGATGGAATAGCCGCG	HindIII
Orf29-1 b-His (MC58)	Fwd	CGCGGATCCGCTAGC-AACGGTTTGATGCCCCG	NheI
	Rev	CCCGCTCGAG-TTTGTCTAAGTTCTGTATAT CCCGCTCGAG-ATTCCACCTGCCATC	XhoI
Orf29-1 b-L (MC58)	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
	Rev	CCCGCTCGAG-TTAATTCCACCTGCCATC	XhoI
Orf29-1 c-His (MC58)	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
	Rev	CCCGCTCGAG-TTGGACGATGCCCCGGA	XhoI
Orf29-1 c-L (MC58)	Fwd	CGCGGATCCGCTAGC-ATGAATTTGCCTATTCAAAAAT	NheI
	Rev	CCCGCTCGAG-TTATTGGACGATGCCCCG	XhoI
Orf25L	Fwd	CGCGGATCCCATATG-TATCGCAAACGTATTGC	NdeI
	Rev	CCCGCTCGAG-CTAATCGATGGAATAGCC	XhoI
Orf37L	Fwd	CGCGGATCCCATATG-AAACAGACAGTCAAATG	NdeI
	Rev	CCCGCTCGAG-TCAATAACCCGCCTTCAG	XhoI
Orf38L	Fwd	CGCGGATCCCATATG- TTACGTTTGACTGCTTTAGCCGTATGCACC	NdeI
	Rev	CCCGCTCGAG- TTATTTTGCCGCGTTAAAGCGTCGGCAAC	XhoI
Orf40L	Fwd	CGCGGATCCCATATG-AACAAAATATACCGCAT	NdeI
	Rev	CCCGCTCGAG-TTACCACTGATAACCGAC	XhoI
Orf40.2-His	Fwd	CGCGGATCCCATATG-ACCGATGACGACGATTAT	NdeI
	Rev	GCCCAAGCTT-CCACTGATAACCGACAGA	HindIII
Orf40.2L	Fwd	CGCGGATCCCATATG-AACAAAATATACCGCAT	NdeI
	Rev	GCCCAAGCTT-TTACCACTGATAACCGAC	HindIII
Orf46-2L	Fwd	GGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI
	Rev	CCCGCTCGAG-TTATTACTCTATAACGAGGTCTCTTAAC	XhoI
Orf46-2	Fwd	GGAATTCCATATG-TCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAG-TTATTACTCTATAACGAGGTCTCTTAAC	XhoI
Orf46.1L	Fwd	GGAATTCCATATG-GGCATTTCCCGCAAAATATC	NdeI

	Rev	CCCGCTCGAG-TTACGTATCATATTTACGTGC	XhoI
orf46. (His-GST)	Fwd	GGGAATTCCATATGCACGTGAAATATGATACGAAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTTACTCCTATAACGAGGTCTCTTAAC	XhoI
orf46.1-His	Fwd	GGGAATTCCATATGTCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGCGTATCATATTTACGTGC	XhoI
orf46.2-His	Fwd	GGGAATTCCATATGTCAGATTTGGCAAACGATTCTT	NdeI
	Rev	CCCGCTCGAGTTTACTCCTATAACGAGGTCTCTTAAC	XhoI
Orf65-1-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-CAAAATGCGTTCAAAATCCC	BamHI-NdeI
	Rev	CGCGGATCCCATATG-AACAAAATATACCGCAT CCCGCTCGAG-TTTGCTTTTCGATAGAACGG	XhoI
Orf72-1L	Fwd	GCGGCCATATG-GTCATAAAATATACAAATTTGAA	NdeI
	Rev	GCGGCCTCGAG-TTAGCCTGAGACCTTTGCAAATT	XhoI
Orf76-1L	Fwd	GCGGCCATATG-AAACAGAAAAAAACCGCTG	NdeI
	Rev	GCGGCCTCGAG-TTACGGTTTGACACCGTTTTTC	XhoI
Orf83-1L	Fwd	CGCGGATCCCATATG-AAAACCTGCTCCTC	NdeI
	Rev	CCCGCTCGAG-TTATCCTCCTTTGCGGC	XhoI
Orf85-2L	Fwd	GCGGCCATATG-GCAAAAATGATGAAATGGG	NdeI
	Rev	GCGGCCTCGAG-TTATCGGCGCGGCGGGCC	XhoI
Orf91L (MC58)	Fwd	GCGGCCATATGAAAAAATCCTCCCTCATCA	NdeI
	Rev	GCGGCCTCGAGTTATTTGCCGCCGTTTTTGGC	XhoI
Orf91-His(MC58)	Fwd	GCGGCCATATGGCCCTGCGGACGCGGTAAG	NdeI
	Rev	GCGGCCTCGAGTTTGCCGCCGTTTTTGGCTTTC	XhoI
Orf97-1L	Fwd	GCGGCCATATG-AAACACATACTCCCCCTGA	NdeI
	Rev	GCGGCCTCGAG-TTATTGCTCTACGGTTTTTTG	XhoI
Orf119L (MC58)	Fwd	GCGGCCATATGATTTACATCGTACTGTTTC	NdeI
	Rev	GCGGCCTCGAGTTAGGAGAACAGGCGCAATGC	XhoI
Orf119-His(MC58)	Fwd	GCGGCCATATGTACAACATGTATCAGGAAAC	NdeI
	Rev	GCGGCCTCGAGGAGAACAGGCGCAATGCGG	XhoI
Orf137.1 (His-GST) (MC58)	Fwd	CGCGGATCCGCTAGCTGCGGCACGGCGGG	BamHI-NheI
	Rev	CCCGCTCGAGATAACGGTATGCCGCCAG	XhoI
Orf143-1L	Fwd	CGCGGATCCCATATG-GAATCAACACTTTCAC	NdeI
	Rev	CCCGCTCGAG-TTACACGCGGTGCTGT	XhoI
008	Fwd	CGCGGATCCCATATG-AACAACAGACATTTTG	NdeI
	Rev	CCCGCTCGAG-TTACCTGTCCGGTAAAAG	XhoI
050-1(48)	Fwd	CGCGGATCCGCTAGC-ACCGTCATCAAACAGGAA	NheI
	Rev	CCCGCTCGAG-TCAAGATTCGACGGGGA	XhoI
105	Fwd	CGCGGATCCCATATG-TCCGCAAACGAATACG	NdeI
	Rev	CCCGCTCGAG-TCAGTGTTCTGCCAGTTT	XhoI
111L	Fwd	CGCGGATCCCATATG-CCGTCTGAAACACG	NdeI
	Rev	CCCGCTCGAG-TTAGCGGAGCAGTTTTTC	XhoI
117-1	Fwd	CGCGGATCCCATATG-ACCGCCATCAGCC	NdeI
	Rev	CCCGCTCGAG-TTAAAGCCGGGTAACGC	XhoI
121-1	Fwd	GCGGCCATATG-GAAACACAGCTTTACATCGG	NdeI
	Rev	GCGGCCTCGAG-TCAATAATAATATCCCGCG	XhoI

122-1	Fwd	GCGGCCATATG-ATTAAAATCCGCAATATCC	NdeI
	Rev	GCGGCCTCGAG-TTAAATCTTGGTAGATTGGATTGG	XhoI
128-1	Fwd	GCGGCCATATG-ACTGACAACGCACTGCTCC	NdeI
	Rev	GCGGCCTCGAG-TCAGACCGCGTTGTGCGAAAC	XhoI
148	Fwd	CGCGGATCCCATATG-GCGTTAAAAACATCAAA	NdeI
	Rev	CCCGCTCGAG-TCAGCCCTTCATACAGC	XhoI
149.1L (MC58)	Fwd	GCGGCATTAATGGCACAACACTACACTCAAACC	AseI
	Rev	GCGGCCTCGAGTTAAAACTTCACGTTACGCGG	XhoI
149.1-His(MC58)	Fwd	GCGGCATTAATGCATGAAACTGAGCAATCGGTGG	AseI
	Rev	GCGGCCTCGAGAACTTCACGTTACGCGCGCGTAAA	XhoI
205 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGGCGAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCGCTCGAGATAATGGCGGCGGCGG	XhoI
206L	Fwd	CGCGGATCCCATATG-TTCCCCCGACAA	NdeI
	Rev	CCCGCTCGAG-TCATTCTGTAAAAAAGTATG	XhoI
214 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGCTTCAAAGCGACAGCAG	BamHI-NdeI
	Rev	CCCGCTCGAGTTCCGATTTTTCGCTACTC	XhoI
216	Fwd	CGCGGATCCCATATG-GCAATGGCAGAAAACG	NdeI
	Rev	CCCGCTCGAG-CTATACAATCCGTGCCG	XhoI
225-1L	Fwd	CGCGGATCCCATATG-GATTCTTTTTTCAAACC	NdeI
	Rev	CCCGCTCGAG-TCAGTTCAGAAAGCGGG	XhoI
235L	Fwd	CGCGGATCCCATATG-AAACCTTTGATTTTAGG	NdeI
	Rev	CCCGCTCGAG-TTATTTGGGCTGCTCTTC	XhoI
243	Fwd	CGCGGATCCCATATG-GTAATCGTCTGGTTG	NdeI
	Rev	CCCGCTCGAG-CTACGACTTGTTACCG	XhoI
247-1L	Fwd	GCGGCCATATG-AGACGTAAAATGCTAAAGCTAC	NdeI
	Rev	GCGGCCTCGAG-TCAAAGTGTTCTGTTGCGC	XhoI
264-His	Fwd	GCCGCCATATG-TTGACTTTAACCCGAAAAA	NdeI
	Rev	GCCGCGTCGAG-GCCGGCGGTCAATACCGCCCGAA	XhoI
270 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGCGCAATGCGATTGAC	BamHI-NdeI
	Rev	CCCGCTCGAGTTCCGCGGTAAATGCCG	XhoI
274L	Fwd	GCGGCCATATG-GCGGGGCCGATTTTTGT	NdeI
	Rev	GCGGCCTCGAG-TTATTTGCTTTCAGTATTATTG	XhoI
283L	Fwd	GCGGCCATATG-AACTTTGCTTTATCCGTCA	NdeI
	Rev	GCGGCCTCGAG-TTAACGGCAGTATTTGTTTAC	XhoI
285-His	Fwd	CGCGGATCCCATATGGGTTTGGCGTTCCGGC	BamHI
	Rev	GCCCAAGCTTTTTTCTTTGCCGTTCCG	HindIII
286-His (MC58)	Fwd	CGCGGATCCCATATG-GCCGACCTTCCGAAAA	NdeI
	Rev	CCCGCTCGAG-GAAGCGCGTTCCCAAGC	XhoI
286L (MC58)	Fwd	CGCGGATCCCATATG-CACGACACCCGTAC	NdeI
	Rev	CCCGCTCGAG-TTAGAAGCGCGTTCCCAA	XhoI
287L	Fwd	CTAGCTAGC-TTAAACGCAGCGTAATCGCAATGG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTGGC	XhoI

287	Fwd	CTAGCTAGC-GGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
287L ^{Orf4}	Fwd	CTAGCTAGCGCTCATCCTCGCCGCC-TGCGGGGGCGGCGGT	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
287-fu	Fwd	CGGGGATCC-GGGGGCGGCGGTGGCG	BamHI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
287-His	Fwd	CTAGCTAGC-GGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC *	XhoI
287-His(2996)	Fwd	CTAGCTAGC-TGCGGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC	XhoI
Δ1 287-His	Fwd	CGCGGATCCGCTAGC-CCCGATGTAAATCGGC [‡]	NheI
Δ2 287-His	Fwd	CGCGGATCCGCTAGC-CAAGATATGGCGCAGT [‡]	NheI
Δ3 287-His	Fwd	CGCGGATCCGCTAGC-GCCGAATCCGCAAATCA [‡]	NheI
Δ4 287-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTGGCTAATGG [‡]	NheI
Δ4 287MC58-His	Fwd	CGCGCTAGC-GGAAGGGTTGATTGGCTAATGG [‡]	NheI
287a-His	Fwd	CGCCATATG-TTTAAACGCAGCGTAATCGC	NdeI
	Rev	CCCGCTCGAG-AAAATTGCTACCGCATTGCGCAGG	XhoI
287b-His	Fwd	CGCCATATG-GGAAGGGTTGATTGGCTAATGG	NdeI
287b-2996-His	Rev	CCCGCTCGAG-CTTGCTTTTATAAATGATGACATATTG	XhoI
287b-MC58-His	Rev	CCCGCTCGAG-TTTATAAAAGATAATATATTGATTGATTCC	XhoI
287c-2996-His	Fwd	CGCGCTAGC-ATGCCGCTGATTCCCGTCAATC [‡]	NheI
‘287 未标记的’(2996)	Fwd	CTAGCTAGC-GGGGGCGGCGGTGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG287-His *	Fwd	CGCGGATCCGCTAGC-CCCGATGTAAATCGGC	NheI
	Rev	CCCGCTCGAG-ATCCTGCTCTTTTTTGCC	XhoI
ΔG287K(2996)	Fwd	CGCGGATCCGCTAGC-CCCGATGTAAATCGGC	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG 287-L	Fwd	CGCGGATCCGCTAGC-TTGAAACGCAGTGTGATTGCAATGGCTTGATTTTTTGCC CTTTCAGCCTGT TCGCCCGATGTAAATCGGCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
ΔG 287-Orf4L	Fwd	CGCGGATCCGCTAGC-AAAACCTTCTTCAAAAGCCTTTCCGCCGCCGCACTCGCG CTCATCCTCGCCGCTGC TCGCCCGATGTAAATCG	NheI
	Rev	CCCGCTCGAG-TCAATCCTGCTCTTTTTTGCC	XhoI
292L	Fwd	CGCGGATCCCATATG-AAAACCAAGTTAATCAAA	NdeI
	Rev	CCCGCTCGAG-TTATTGATTTTTGCGATGA	XhoI
308-1	Fwd	CGCGGATCCCATATG-TTAAATCGGGTATTTTATC	NdeI
	Rev	CCCGCTCGAG-TTAATCCGCCATTCCCTG	XhoI
401L	Fwd	GCGGCCATATG-AAATTACAACAATTGGCTG	NdeI
	Rev	GCGGCCTCGAG-TTACCTTACGTTTTTCAAAG	XhoI
406L	Fwd	CGCGGATCCCATATG-CAAGCACGGCTGCT	NdeI
	Rev	CCCGCTCGAG-TCAAGGTTGTCCTTGTCTA	XhoI
502-1L	Fwd	CGCGGATCCCATATG-ATGAAACCGCACAAAC	NdeI
	Rev	CCCGCTCGAG-TCAAGTTGCTCAACACGTC	XhoI

502-A (His-GST)	Fwd	CGCGGATCCCATATG GTAGACGCGCTTAAGCA	BamHI-NdeI
	Rev	CCCGCTCGAGAGCTGCATGGCGGCG	XhoI
503-1L	Fwd	CGCGGATCCCATATG GCACGGTCGTTATAC	NdeI
	Rev	CCCGCTCGAG-CTACCGCGCATTCTG	XhoI
519-1L	Fwd	GCGGCCATATG-GAATTTTTCATTATCTTGT	NdeI
	Rev	GCGGCCCTCGAG-TTATTTGGCGGTTTGTCTGC	XhoI
525-1L	Fwd	GCGGCCCATATG-AAGTATGTCCGGTTATTTTC	NdeI
	Rev	GCGGCCCTCGAG-TTATCGGCTTGTGCAACGG	XhoI
529-(His/GST) (MC58)	Fwd	CGCGGATCCGCTAGC-TCCGGCAGCAAACCGA	Bam HI-NheI
	Rev	GCCCAAGCTT-ACGCAGTTCGGAATGGAG	HindIII
552L	Fwd	GCCGCCATATGTTGAATATTAACTGAAAACCTTG	NdeI
	Rev	GCCGCCTCGAGTTATTTCTGATGCCTTTCCC	XhoI
556L	Fwd	GCCGCCATATGGACAATAAGACCAAACCTG	NdeI
	Rev	GCCGCCTCGAGTTAACGGTGC GGACGTTTC	XhoI
557L	Fwd	CGCGGATCCCATATG-AACAAACTGTTTCTTAC	NdeI
	Rev	CCCGCTCGAG-TCATTCCGCCTTCAGAAA	XhoI
564ab-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG- CAAGGTATCGTTGCCGACAAATCCGCACCT	BamHI-NdeI
	Rev	CCCGCTCGAG- AGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	XhoI
564abL (MC58)	Fwd	CGCGGATCCCATATG- AACCGCACCCGTACAAAAGTTGTATTAAACAAACATC	NdeI
	Rev	CCCGCTCGAG- TTAAGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	XhoI
564b- (His/GST)(MC58)	Fwd	CGCGGATCCCATATG- ACGGGAGAAAATCATGCGGTTTCACTTCATG	BamHI-NdeI
	Rev	CCCGCTCGAG- AGCTAATTGTGCTTGGTTTGCAGATAGGAGTT	XhoI
564c- (His/GST)(MC58)	Fwd	CGCGGATCCCATATG- GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	BamHI-NdeI
	Rev	CCCGCTCGAG- GCGGTAAC TGCCGCTTGCACTGAATCCGTAA	XhoI
564bc- (His/GST)(MC58)	Fwd	CGCGGATCCCATATG- ACGGGAGAAAATCATGCGGTTTCACTTCATG	BamHI-NdeI
	Rev	CCCGCTCGAG- GCGGTAAC TGCCGCTTGCACTGAATCCGTAA	XhoI
564d- (His/GST)(MC58)	Fwd	CGCGGATCCCATATG- CAAAGCAAAGTCAAAGCAGACCATGCCTCCGTAA	BamHI-NdeI
	Rev	CCCGCTCGAG- TCITTTCTTTCAATTATAACTTTAGTAGGTTCAATTTTG GTCCCC	XhoI
564cd- (His/GST)(MC58)	Fwd	CGCGGATCCCATATG- GTTTCAGACGGCCTATACAACCAACATGGTGAAATT	BamHI-NdeI
	Rev	CCCGCTCGAG- TCITTTCTTTCAATTATAACTTTAGTAGGTTCAATTTTG GTCCCC	XhoI
570L	Fwd	GCGGCCATATG-ACCCGTTTGACCCGCG	NdeI
	Rev	GCGGCCCTCGAG-TCAGCGGGCGTTCATTTCTT	XhoI
576-1L	Fwd	CGCGGATCCCATATG-AACACCATTTTCAAATC	NdeI
	Rev	CCCGCTCGAG-TTAATTTACTTTTTTGATGTGCG	XhoI

580L	Fwd	GCGGCCATATG-GATTCGCCCAAGGTCGG	NdeI
	Rev	GCGGCCTCGAG-CTACACTTCCCCGAAGTGG	XhoI
583L	Fwd	CGCGGATCCCATATG-ATAGTTGACCAAAGCC	NdeI
	Rev	CCCGCTCGAG-TTATTTTCCGATTTTTCGG	XhoI
593	Fwd	GCGGCCATATG-CTTGAAGTGAACGGACT	NdeI
	Rev	GCGGCCTCGAG-TCAGCGGAAGCGGACGATT	XhoI
650 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGTCCAAACTCAAACCATCG	BamHI-NdeI
	Rev	CCCGCTCGAGGCTTCCAATCAGTTTGACC	XhoI
652	Fwd	GCGGCCATATG-AGCGCAATCGTTGATATTTTC	NdeI
	Rev	GCGGCCTCGAG-TTATTTGCCAGTTGGTAGAATG	XhoI
664L	Fwd	GCGGCCATATG-GTGATACATCCGCACTACTTC	NdeI
	Rev	GCGGCCTCGAG-TCAAAATCGAGTTTACACCA	XhoI
726	Fwd	GCGGCCATATG-ACCATCTATTTCAAAAACGG	NdeI
	Rev	GCGGCCTCGAG-TCAGCCGATGTTTAGCGTCCATT	XhoI
741-His(MC58)	Fwd	CGCGGATCCCATATG-AGCAGCGGAGGGGGTG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGCGCGCAAGGC	XhoI
ΔG741-His(MC58)	Fwd	CGCGGATCCCATATG-GTCGCCGCCGACATCG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGCGCGCAAGGC	XhoI
686-2-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-OGCGGTTCCGAAGGCG	BamHI-NdeI
	Rev	CCCGCTCGAG-TTGAACACTGATGTCTTTCCGA	XhoI
719-(His/GST) (MC58)	Fwd	CGCGGATCCGCTAGC-AACTGTCTGTTGGTGTTAAC	BamHI-NheI
	Rev	CCCGCTCGAG-TTGACCCGCTCCACGG	XhoI
730-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGCGCAAGACCC	NdeI
	Rev	GCCGCCTCGAGATCTCTAAACCTGTTTAAACAATGCCG	XhoI
730A-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGCGCAAGACCC	NdeI
	Rev	GCGGCCTCGAGCTCCATGCTGTTGCCCCAGC	XhoI
730B-His (MC58)	Fwd	GCCGCCATATGGCGGACTTGCGCAAGACCC	NdeI
	Rev	GCGGCCTCGAGAAAAATCCCCGCTAACCCGAG	XhoI
741-His (MC58)	Fwd	CGCGGATCCCATATG-AGCAGCGGAGGGGGTG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGCGCGCAAGGC	XhoI
ΔG741-His (MC58)	Fwd	CGCGGATCCCATATG-GTCGCCGCCGACATCG	NdeI
	Rev	CCCGCTCGAG-TTGCTTGCGCGCAAGGC	XhoI
743 (His-GST)	Fwd	CGCGGATCCCATATGGACGGTGTGTGCCTGTT	BamHI-NdeI
	Rev	CCCGCTCGAGCTTACGGATCAAATTGACG	XhoI
757 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGGGCAGCCAATCTGAAGAA	BamHI-NdeI
	Rev	CCCGCTCGAGCTCAGCTTTTGCCGTCAA	XhoI
759-His/GST (MC58)	Fwd	CGCGGATCCGCTAGC-TACTCATCCATTGTCCGC	BamHI-NheI
	Rev	CCCGCTCGAG-CCAGTTGTAGCCTATTTG	XhoI
759L (MC58)	Fwd	CGCGGATCCGCTAGC-ATGCGCTTACACACAC	NheI
	Rev	CCCGCTCGAG-TTACCAGTTGTAGCCTATTT	XhoI
760-His	Fwd	GCCGCCATATGGCACAAACGGAAGGTTTGAA	NdeI
	Rev	GCCGCCTCGAGAAAACTGTAACGCAGGTTTGCCGTC	XhoI
769-His (MC58)	Fwd	GCGGCCATATGGAAGAAACACCGCGCAACCG	NdeI

	Rev	GCGGCCTCGAGGAACGTTTTATTAAACTCGAC	XhoI
907L	Fwd	GCGGCCATATG-AGAAAACCGACCGATACCCTA	NdeI
	Rev	GCGGCCTCGAG-TCAACGCCACTGCCAGCGGTTG	XhoI
911L	Fwd	COCOOATCCCATATG-AAGAAAGACATATTGGAATTTGGGTCGGACTG	NdeI
	Rev	CCCGCTCGAG-TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
911LOmpA	Fwd	GGGAATTCATATGAAAAAGACAGCTATCGCGATTGCA GTGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCCGC TAGC-GCTTTCGCGGTGGCCGCGCGTGC	NdeI-(NheI)
	Rev	CCCGCTCGAG-TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
911LPelB	Fwd	CATGCCATGG-CTTCCGCGTGGCCGGCGGTGC	NcoI
	Rev	CCCGCTCGAG-TTATTCGGCGGCTTTTTCCGCATTGCCG	XhoI
913-His/GST (MC58)	Fwd	CGCGGATCCCATATG-TTGCCGAAACCCGCC	BamHI-NdeI
	Rev	CCCGCTCGAG-AGGTTGTGTTCCAGGTTG	XhoI
913L (MC58)	Fwd	CGCGGATCCCATATG-AAAAAAACCGCCTATG	NdeI
	Rev	CCCGCTCGAG-TTAAGGTTGTGTTCCAGG	XhoI
919L	Fwd	CGCGGATCCCATATG-AAAAAATACCTATTCCGC	NdeI
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCCG	XhoI
919	Fwd	CGCGGATCCCATATG-CAAAGCAAGAGCATCCAAA	NdeI
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCCG	XhoI
919L Orf4	Fwd	GGGAATTCATATGAAAACCTTCTTCAAACCCCTTCCG CCGCCGCTAGCGCTCATCTCGCCGCC- TGCCAAAGCAAGAGCATC	NdeI-(NheI)
	Rev	CCCGCTCGAG-TTACGGGCGGTATTCCGGCTTCATACCG	XhoI
(919)-287fusion	Fwd	CGCGGATCCGTCGAC-TGTGGGGCGGCGGTGGC	Sall
	Rev	CCCGCTCGAG-TCAATCCGCTCTTTTTTGCC	XhoI
920-1L	Fwd	GCGGCCATATG-AAGAAAACATTGACACTGC	NdeI
	Rev	GCGGCCTCGAG-TTAATGGTGCGAATGACCGAT	XhoI
925-His/GST (MC58) _{GATE}	Fwd	ggggacaagttgtacaaaaagcaggcTGC GGCAAGGATGCCGG	attB1
	Rev	ggggaccactttgtacaagaaagctgggCTAAAGCAACAATGCCGG	attB2
926L	Fwd	CGCGGATCCCATATG-AAACACACCGTATCC	NdeI
	Rev	CCCGCTCGAG-TTATCTCGTGC GCGCC	XhoI
927-2-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-AGCCCCGCGCCGATT	BamHI-NdeI
	Rev	CCCGCTCGAG-TTTTTGTGCGGT CAGGCG	XhoI
932-His/GST (MC58) _{GATE}	Fwd	ggggacaagttgtacaaaaagcaggcTGTTCGTTTGGGGATT TAA ACCAAACCAAATC	attB1
935 (His-GST) (MC58)	For	CGCGGATCCCATATGGCGGATGCGCCCGCG	BamHI-NdeI
	Rev	CCCGCTCGAGAAACCGCCAATCCGCC	XhoI
936-1L	Rev	ggggaccactttgtacaagaaagctgggTCATTTTGT TTTTCCTTCTTCT CGAGGCCATT	attB2
	Fwd	CGCGGATCCCATATG-AAACCCAAACCGCAC	NdeI
	Rev	CCCGCTCGAG-TCAGCGTTGGACGTAGT	XhoI
953L	Fwd	GGGAATTCATATG-AAAAAATCATCTTCGCCG	NdeI
	Rev	CCCGCTCGAG-TTATTGTTTGGCTGCCTCGAT	XhoI
953-fu	Fwd	GGGAATTCATATG-GCCACCTACAAAGTGGACG	NdeI
	Rev	CGGGGATCC-TTGTTTGGCTGCCTCGATTG	BamHI

954 (His-GST) (MC58)	Fwd	CGCGGATCCCATATGCAAGAACAATCGCAGAAAG	BamHI-NdeI
	Rev	CCCCTCGAGTTTTTTCGGCAAATTGGCTT	XhoI
958-His/GST (MC58) ^{GATE}	Fwd	ggggacaagttgtacaaaaagcaggctGCCGATGCCGTTGCGG	attB1
961L	Rev	ggggaccacttgtacaagaagctgggtTCAGGGTCGTTTGTGTCG	attB2
	Fwd	CGCGGATCCCATATG-AAACACTTTCCATCC	NdeI
	Rev	CCCCTCGAG-TTACCACTCGTAATTGAC	XhoI
961	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGAC	NdeI
	Rev	CCCCTCGAG-TTACCACTCGTAATTGAC	XhoI
961 c (His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	BamHI-NdeI
	Rev	CCCCTCGAG-ACCCACGTTGTAAGGTTG	XhoI
961 c-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACGA	BamHI-NdeI
	Rev	CCCCTCGAG-ACCCACGTTGTAAGGTTG	XhoI
961 c-L	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961 c-L (MC58)	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCCTCGAG-TTAACCCACGTTGTAAGGT	XhoI
961 d (His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACG	BamHI-NdeI
	Rev	CCCCTCGAG-GTCTGACACTGTTTATCC	XhoI
961 Δ1-L	Fwd	CGCGGATCCCATATG-ATGAAACACTTTCCATCC	NdeI
	Rev	CCCCTCGAG-TTATGCTTTGGCGGCAAAG	XhoI
fu 961-...	Fwd	CGCGGATCCCATATG- GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961-... (MC58)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGAC	NdeI
	Rev	CGCGGATCC-CCACTCGTAATTGACGCC	BamHI
fu 961 c -...	Fwd	CGCGGATCCCATATG-GCCACAAACGACGAC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu 961 c-L-...	Fwd	CGCGGATCCCATATG- ATGAAACACTTTCCATCC	NdeI
	Rev	CGCGGATCC -ACCCACGTTGTAAGGTTG	BamHI
fu (961)- 741(MC58)-His	Fwd	CGCGGATCC -GGAGGGGGTGGTGTCG	BamHI
	Rev	CCCCTCGAG-TTGCTTGGCGGCAAGGC	XhoI
fu (961)-983-His	Fwd	CGCGGATCC - GGCGAGGCGGCACTT	BamHI
	Rev	CCCCTCGAG-GAACCGGTAGCCTACG	XhoI
fu (961)- Orf46.1- His	Fwd	CGCGGATCCGGTGGTGGTGGT- TCAGATTGGCAAACGATTC	BamHI
	Rev	CCCCTCGAG-CGTATCATATTTACGTGC	XhoI
fu (961 c-L)- 741(MC58)	Fwd	CGCGGATCC -GGAGGGGGTGGTGTCG	BamHI
	Rev	CCCCTCGAG-TTATTGCTTGGCGGCAAG	XhoI
fu (961c-L)-983	Fwd	CGCGGATCC - GGCGAGGCGGCACTT	BamHI
	Rev	CCCCTCGAG-TCAGAACCGGTAGCCTAC	XhoI
fu (961c-L)- Orf46.1	Fwd	CGCGGATCCGGTGGTGGTGGT- TCAGATTGGCAAACGATTC	BamHI
	Rev	CCCCTCGAG-TTACGTATCATATTTACGTGC	XhoI
961-(His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAGCGACGACG	BamHI-NdeI

(MC58)	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI
961 Δ1-His	Fwd	CGCGGATCCCATATG-GCCACAAACGACGAC	NdeI
	Rev	CCCGCTCGAG-TGCTTTGGCGGCAAAGTT	XhoI
961a-(His/GST)	Fwd	CGCGGATCCCATATG-GCCACAAACGACGAC	BamHI-NdeI
	Rev	CCCGCTCGAG-TTAGCAATATTATCTTTGTTCTGTAGC	XhoI
961b-(His/GST)	Fwd	CGCGGATCCCATATG-AAAGCAAACCGTGCCGA	BamHI-NdeI
	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI
961-His/GST ^{GATE}	Fwd	ggggacaagttgtacaaaaagcaggctGCAGCCACAAACGACGACG ATGTTAAAAAAGC	attB1
	Rev	ggggaccactttgtacaagaagctgggtTTACCACTCGTAATTGACGC CGACATGGTAGG	attB2
982	Fwd	GCGGCCATATG-GCAGCAAAAGACGTACAGTT	NdeI
	Rev	GCGGCCCTCGAG-TTACATCATGCCGCCCATACCA	XhoI
983-His (2996)	Fwd	CGCGGATCCGCTAGC-TTAGGCGGCGGCGGAG	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
ΔG983-His (2996)	Fwd	CCCCTAGCTAGC-ACCTCTGCGCCGACTT	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
983-His	Fwd	CGCGGATCCGCTAGC-TTAGGCGGCGGCGGAG	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
ΔG983-His	Fwd	CGCGGATCCGCTAGC-ACCTCTGCGCCGACTT	NheI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
983L	Fwd	CGCGGATCCGCTAGC- CGAACGACCCCAACCTTCCCTACAAAACTTTCAA	NheI
	Rev	CCCGCTCGAG-TCAGAACCGACGTGCCAAGCCGTTT	XhoI
987-His (MC58)	Fwd	GCCGCCATATGCCCCCACTGGAAGAACGGACG	NdeI
	Rev	GCCGCCCTCGAGTAATAAACCTTCTATGGGCAGCAG	XhoI
989-(His/GST) (MC58)	Fwd	CGCGGATCCCATATG-TCCGTCCACGCATCCG	BamHI-NdeI
	Rev	CCCGCTCGAG-TTTGAATTTGTAGGTGTATTG	XhoI
989L (MC58)	Fwd	CGCGGATCCCATATG-ACCCCTTCCGCACT	NdeI
	Rev	CCCGCTCGAG-TTATTTGAATTTGTAGGTGTAT	XhoI
CrgA-His (MC58)	Fwd	CGCGGATCCCATATG-AAACCAATTTCAGAAGAA	NdeI
	Rev	CCCGCTCGAG-TCCACAGAGATTGTTTCC	XhoI
PlIC1-ES (MC58)	Fwd	GATGCCCGAAGGGCGGG	
	Rev	GCCCAAGCTT-TCAGAAGAAGACTTCACGC	
PlIC1-His (MC58)	Fwd	CGCGGATCCCATATG-CAAAACCATAAATACGCTATT	NdeI
	Rev	GCCCAAGCTT-GAAGAAGACTTCACGCCAG	HindIII
Δ1PlIC1-His (MC58)	Fwd	CGCGGATCCCATATG-GTCTTTTTCGACAATACCGA	NdeI
	Rev	GCCCAAGCTT-	HindIII
PlIC1L (MC58)	Fwd	CGCGGATCCCATATG-AATAAACTTTAAAAAGGCGG	NdeI
	Rev	GCCCAAGCTT-TCAGAAGAAGACTTCACGC	HindIII
ΔGTbp2-His (MC58)	Fwd	CGCGAATCCCATATG-TTCGATCTTGATTCTGTGCA	NdeI
	Rev	CCCGCTCGAG-TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His (MC58)	Fwd	CGCGAATCCCATATG-TTGGGCGGAGGCGGCAG	NdeI
	Rev	CCCGCTCGAG-TCGCACAGGCTGTTGGCG	XhoI
Tbp2-His(MC58)	Fwd	CGCGAATCCCATATG-TTGGGCGGAGGCGGCAG	NdeI
	Rev	CCCGCTCGAG-TCGCACAGGCTGTTGGCG	XhoI

*

NMB0109- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GCAAATTTGGAGGTGCGC	BamHI-NdeI
	Rev	CCCCTCGAG-TTCGGAGCGGTTGAAGC	XhoI
NMB0109L (MC58)	Fwd	CGCGGATCCCATATG-CAACGTCGTATTATAACCC	NdeI
	Rev	CCCCTCGAG-TTATTCGGAGCGGTTGAAG	XhoI
NMB0207- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG- GGCATCAAAGTCGCCATCAACGGCTAC	BamHI-NdeI
	Rev	CCCCTCGAG-TTTGAGCGGGCGCACTTCAAGTCCG	XhoI
NMB0462- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GGCGGCAGCGAAAAAAC	BamHI-NdeI
	Rev	CCCCTCGAG-GTTGGTGCCGACTTTGAT	XhoI
NMB0623- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-GGCGGCGGAAGCGATA	BamHI-NdeI
	Rev	CCCCTCGAG-TTTGCCCGCTTTGAGCC	XhoI
NMB0625 (His- GST)(MC58)	Fwd	CGCGGATCCCATATGGGCCAAATCCGAAAATACG	BamHI-NdeI
	Rev	CCCCTCGAGCATCCCGTACTGTTTCG	XhoI
NMB0634 (His/GST)(MC58)	Fwd	ggggacaagttgtacaaaaagcaggctCCGACATTACCGTGTAACAAC GGCCAACAAAGAA	attB1
	Rev	ggggaccactttgtacaagaagctgggtCTTATTTTCATACCGGCTTGCT CAAGCAGCCGG	attB2
NMB0776- His/GST (MC58) GATE	Fwd	ggggacaagttgtacaaaaagcaggctGATACGGTGTTTTCTGTAA AACGGACAACAA	attB1
	Rev	ggggaccactttgtacaagaagctgggtCTAGGAAAAATCGTCATCGT TGAAATTCGCC	attB2
NMB1115- His/GST (MC58) GATE	Fwd	ggggacaagttgtacaaaaagcaggctATGCACCCCATCGAAACC	attB1
	Rev	ggggaccactttgtacaagaagctgggtCTAGTCTTGCAGTGCGCTC	attB2
NMB1343- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG- GGAAATTTCTTATATAGAGGCATTAG	BamHI-NdeI
	Rev	CCCCTCGAG- GTAAATTTCTATCAACTCTTTAGCAATAAT	XhoI
NMB1369 (His- GST) (MC58)	Fwd	CGCGGATCCCATATGGCCTGCCAAGACGACA	BamHI-NdeI
	Rev	CCCCTCGAGCGCCTCCTGCCGAAA	XhoI
NMB1551 (His- GST)(MC58)	Fwd	CGCGGATCCCATATGGCAGAGATCTGTTTGATAA	BamHI-NdeI
	Rev	CCCCTCGAGCGGTTTTCCGCCCAATG	XhoI
NMB1899 (His- GST) (MC58)	Fwd	CGCGGATCCCATATGCAGCCGGATACGGTC	BamHI-NdeI
	Rev	CCCCTCGAGAATCACTTCCAACACAAAAT	XhoI
NMB2050- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG-TGGTTGCTGATGAAGGGC	BamHI-NdeI
	Rev	CCCCTCGAG-GACTGCTTCATCTTCTGC	XhoI
NMB2050L (MC58)	Fwd	CGCGGATCCCATATG-GAACTGATGACTGTTTTGC	NdeI
	Rev	CCCCTCGAG-TCAGACTGCTTCATCTTCT	XhoI
NMB2159- (His/GST) (MC58)	Fwd	CGCGGATCCCATATG- AGCATTAAAGTAGCGATTAACGGTTTCGGC	BamHI-NdeI
	Rev	CCCCTCGAG- GATTTTGCTGCGAAGTATTCCAAAGTGCG	XhoI
fu-ΔG287....His	Fwd	CGCGGATCCGCTAGC-CCCGATGTTAAATCGGC	NheI

	Rev	CGGGGATCC-ATCCTGCTCTTTTTTGCCGG	BamHI
fu-(ΔG287)-919-His	Fwd	CGCGGATCCGGTGGTGGTGGT-CAAAGCAAGAGCATCCAAACC	BamHI
	Rev	CCCAAGCTT-TTCGGGCGGTATTCGGGCTTC	HindIII
fu-(ΔG287)-953-His	Fwd	CGCGGATCCGGTGGTGGTGGT-GCCACCTACAAAGTGGAC	BamHI
	Rev	GCCCAAGCTT-TTGTTCGGCTGCCTCGAT	HindIII
fu-(ΔG287)-961-His	Fwd	CGCGGATCCGGTGGTGGTGGT-ACAAGCGACGACG	BamHI
	Rev	GCCCAAGCTT-CCACTCGTAATTGACGCC	HindIII
fu-(ΔG287)-Orf46.1-His	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCCAAGCTT-CGTATCATATTTACGTGC	HindIII
fu-(ΔG287)-919-Orf46.1-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	HindIII
	Rev	CCCGCTCGAG-CGTATCATATTTACGTGC	XhoI
fu-(ΔG287-Orf46.1)-919-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGT-CAAAGCAAGAGCATCCAAACC	HindIII
	Rev	CCCGCTCGAG-CGGGCGGTATTCGGGCTT	XhoI
fu ΔG287(394.98)-...	Fwd	CGCGGATCCGCTAGC-CCCAGTGTAAATCGGC	NheI
	Rev	CGGGGATCC-ATCCTGCTCTTTTTTGCCGG	BamHI
fu Orf1-(Orf46.1)-His	Fwd	CGCGGATCCGCTAGC-GGACACACTTATTTCCGGCATC	NheI
	Rev	CGCGGATCC-CCAGCGGTAGCCTAATTTGAT	
fu (Orf1)-Orf46.1-His	Fwd	CGCGGATCCGGTGGTGGTGGT-TCAGATTTGGCAAACGATTC	BamHI
	Rev	CCCAAGCTT-CGTATCATATTTACGTGC	HindIII
fu (919)-Orf46.1-His	Fwd1	GCGGCGTCGACGGTGGCGGAGGCACTGGATCCTCAG	Sall
	Fwd2	GGAGGCACTGGATCCTCAGATTTGGCAAACGATTC	
	Rev	CCCGCTCGAG-CGTATCATATTTACGTGC	XhoI
Fu orf46-....	Fwd	GGAATTCATATGTCAGATTTGGCAAACGATTC	NdeI
	Rev	CGCGGATCCCGTATCATATTTACGTGC	BamHI
Fu (orf46)-287-His	Fwd	CGGGGATCCGGGGCGGGCGGTGGCG	BamHI
	Rev	CCCAAGCTTATCCTGCTCTTTTTTGCCGGC	HindIII
Fu (orf46)-919-His	Fwd	CGCGGATCCGGTGGTGGTGGTCAAAGCAAGAGCATCCAACC	BamHI
	Rev	CCCAAGCTTCGGGCGGTATTCGGGCTTC	HindIII
Fu (orf46-919)-287-His	Fwd	CCCAAGCTTGGGGGCGGGCGGTGGCG	HindIII
	Rev	CCCGCTCGAGATCCTGCTCTTTTTTGCCGGC	XhoI
Fu (orf46-287)-919-His	Fwd	CCCAAGCTTGGTGGTGGTGGTGGTCAAAGCAAGAGCATCCAAACC	HindIII
	Rev	CCCGCTCGAGCGGGCGGTATTCGGGCTT	XhoI
(ΔG741)-961c-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-ACCCAGCTTGTAAGGTTG	XhoI
(ΔG741)-961-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-CCACTCGTAATTGACGCC	XhoI

(ΔG741)-983-His	Fwd	GCGGCCTCGAG-GGATCCGGCGGAGGCGGCACTTCTGCG	XhoI
	Rev	CCCGCTCGAG-GAACCGGTAGCCTACG	XhoI
(ΔG741)-orf46.1-His	Fwd1	GGAGGCACTGGATCCTCAGATTGCGAAACGATTC	SalI
	Fwd2	GCGGCCTCGAGCGGTGGCGGAGGCACTGGATCCTCAGA	
	Rev	CCCGCTCGAG-CGTATCATATTTACGTGC	XhoI
(ΔG983)-741(MC58)-His	Fwd	GCGGCCTCGAG-GGATCCGGAGGGGGTGGTGTGCGCC	XhoI
	Rev	CCCGCTCGAG-TTGCTTGGCGGCAAG	XhoI
(ΔG983)-961c-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-ACCCAGCTTGTAAGGTTG	XhoI
(ΔG983)-961-His	Fwd1	GGAGGCACTGGATCCGCAGCCACAAACGACGACGA	XhoI
	Fwd2	GCGGCCTCGAG-GGTGGCGGAGGCACTGGATCCGCAG	
	Rev	CCCGCTCGAG-CCACTCGTAATTGACGOC	XhoI
(ΔG983)-Orf46.1-His	Fwd1	GGAGGCACTGGATCCTCAGATTGCGAAACGATTC	SalI
	Fwd2	GCGGCCTCGAGCGGTGGCGGAGGCACTGGATCCTCAGA	
	Rev	CCCGCTCGAG-CGTATCATATTTACGTGC	XhoI

*将该引物用作反向引物，将所有 287 的 C 末端融合于 His-标记。

§ 与 287-His 反向引物联用的正向引物。

NB-所有 PCR 反应使用菌株 2996，除非特别指出(如菌株 MC58)。

- 5 在所有以 ATG 起始不跟随唯一 NheI 位点的构建物中，ATG 密码子是用于克隆的 NdeI 位点的一部分。在 5'端用 NheI 作为克隆位点制备的构建物(如所有那些在 N-末端包含 287 的)另有两个融合于抗原的编码序列的密码子(GCT AGC)。

10 染色体 DNA 模板的制备

- 在 100ml GC 培养基中使脑膜炎奈瑟球菌菌株 2996、MC58、394.98、1000 和 BZ232(及其它)生长至指数期，离心收获，并重悬浮于 5ml 缓冲液(20% w/v 蔗糖、50mM Tris-HCl、50mM EDTA、pH 8)中。在冰上培育 10 分钟后，加入 10ml 裂解液(50mM NaCl、1%十二烷基肌氨酸钠(Na-Sarkosyl)、50μg/ml 蛋白酶 K)裂解细菌，悬浮液在 37℃培育 2 小时。进行 2 次苯酚提取(平衡至 pH8)和一次 CHCl₃/异戊醇(24: 1)提取。加入 0.3M 乙酸钠和 2 体积乙醇沉降 DNA，并离心收集。用 70%(v/v)乙醇洗涤沉淀物 1 次，并重溶解于 4.0ml TE 缓冲液(10mM Tris-HCl、1mM EDTA、pH 8.0)。读取 OD₂₆₀ 测定 DNA 浓度。

20 PCR 扩增

- 标准的 PCR 过程进行如下：在 40 μ M 各寡核苷酸引物、400-800 μ M dNTP 溶液、1x PCR 缓冲液(包含 1.5mM MgCl₂)、2.5 单位 TaqI DNA 聚合酶(用 Perkin-Elmer AmpliTaq, Boehringer Mannheim Expand™ 长模板)存在时，将 200ng 2996、MC581000、或 BZ232 菌株的基因组 DNA 或 10ng 重组克隆的质粒 DNA 制备物用作模板。
- 5

- 将整个混合物在 95℃初步培育 3 分钟后，每份样品进行 2-步的扩增：用除去引物(T_{m1})的限制酶尾部的杂交温度进行前 5 轮。然后以按全长？聚物(T_{m2})计算的杂交温度进行 30 轮。根据要扩增的 Orf 的长度，在 68℃或 72℃进行的延伸时间各不相同。对 Orf1 而言，自 3 分钟开始的延伸时间每轮递增 15 秒。
- 10 以在 72℃10 分钟延伸步骤完成循环。

将扩增的 DNA 直接加到 1%琼脂糖凝胶上。按制造商说明，用 Qiagen 凝胶提取试剂盒纯化相应于正确大小条带的 DNA 片段。

PCR 片段和克隆载体的消化

- 15 用适合的限制酶消化相应于扩增片段的纯化 DNA，从而克隆入 pET-21b+、pET22b+、或 pET-24b+。用 QIAquick PCR 纯化试剂盒(按制造商的说明)纯化消化的片段，用 H₂O 或 10mM Tris(pH 8.5)洗脱。用适合的限制酶消化质粒载体，加到 1.0%琼脂糖凝胶上，用 Qiagen A1Qquick 凝胶提取试剂盒纯化相应于消化的载体的条带。

20

克隆

- 将预先消化和纯化的、相应于各基因的片段连接到 pET21b+、pET22b+或 pET-24b+中。将在连接缓冲液(由制造商提供)中的 T4 DNA 连接酶用于摩尔比为 3: 1 的片段/载体。
- 25 通过在冰上培育连接酶反应溶液和细菌 40 分钟，然后在 37℃培育 3 分钟，将重组质粒转化入感受态大肠杆菌 DH5 或 HB101 中。

然后添加 800 μ l LB 肉汤，并在 37℃培育 20 分钟。在 Eppendorf 微型离心机中以最高速度离心这些细胞，并重悬浮于约 200 μ l 的上清液中，并涂布在到 LB 氨苄青霉素(100mg/ml)琼脂上。

在 4.0ml LB 肉汤+100 μ g/ml 氨苄青霉素中培育随机选择的集落过夜，进行重组集落的筛选。使细胞沉淀，并按制造商的说明用 Qiagen QIAprep Spin Miniprep 试剂盒提取质粒 DNA。用适合的限制酶消化约 1 μ g 的各微型制备物，将消化物加到 1-1.5% 琼脂糖凝胶(取决于预计的插入大小)与分子量标记(1kb DNA Ladder, GIBCO)平行。根据插入的大小选择阳性克隆。

表达

各基因克隆入表达载体后，将重组质粒转化入适合表达重组蛋白的大肠杆菌菌株中。如上所述，用 1 μ l 各构建物转化大肠杆菌 BL21-DE3。将单重组集落接种入 2ml LB+Amp(100 μ g/ml)中，在 37 $^{\circ}$ C 培育过夜，然后在 100ml 烧瓶中使用 20ml LB+Amp(100 μ g/ml)以 1: 30 稀释，使 OD₆₀₀ 在 0.1-0.2 之间。将烧瓶置于旋转式水浴摇床中于 30 $^{\circ}$ C 或 37 $^{\circ}$ C 培育，直到 OD₆₀₀ 显示适合诱导表达的指数生长期(0.4-0.8OD)。加入 1.0mM IPTG 诱导蛋白质表达。在 30 $^{\circ}$ C 或 37 $^{\circ}$ C 培育 3 小时后，测定 OD₆₀₀ 并检测表达。用微型离心机离心 1.0ml 各样品，将沉淀物重悬浮于 PBS 中，用 SDS-PAGE 和考马斯蓝染色分析。

Gateway 克隆和表达

克隆 GATE 标记的序列，并用 GATEWAY 克隆方法(GIBCO-BRL)表达。重组克隆(CR)是根据介噬菌体整合入大肠杆菌基因组和从大肠杆菌基因组切除噬菌体的重组反应。整合包括在位于细菌基因组 attB 位点中的噬菌体 DNA 的 attP 位点的重组(BP 反应)和产生侧翼于 attL 和 attR 位点的整合的噬菌体基因组。切除将 attL 和 attR 位点重组回 attP 和 attB 位点(LR 反应)。整合反应需要两种酶[噬菌体蛋白融合酶(Int)和细菌蛋白整合宿主因子(IHF)](BP 克隆酶)。切除反应需要 Int、IHF 和其它噬菌体酶切除酶(Xis)(LR 克隆酶)。将 25-bp 细菌 attB 重组位点的人工衍生物(涉及 B1 和 B2)加到用于 PCR 反应的引物的 5'端，以扩增奈瑟球菌的 ORF。用 BP 克隆酶，将得到的产物 BP 克隆入包含噬菌体 attP 重组位点的互补衍生物的“供体载体”中。得到的“进入克隆”包含侧翼于 attL 位点的衍生物(L1 和 L2)的 ORF，用 LR 克隆酶将其亚克隆入包含 attL-相容的 attR 位点的衍生物的表达“目的载体”中。得到“表达克隆”，其中

ORF 侧翼于 B1 和 B2，且在读框内融合于 GST 或 His N 末端标记。

用于 GATEWAY 表达的大肠杆菌菌株是 BL21-SI。通过在含有盐(0.3M NaCl)的培养基中培育，诱导该菌株的细胞表达 T7 RNA 聚合酶。

要注意的是该系统产生 N-末端 His 标记。

5

膜蛋白的制备

分离主要由内膜、外膜或全膜构成的组分，以获得由膜定位的前导序列表达的重组蛋白。制备膜组分(富含重组蛋白)的方法是采用 Filip 等[J.Bact.(1973) 115:717-722]和 Davies 等人[J.InnunoI.Meth.(1990)143:215-225]的方法。于 37℃，
10 在 20ml LB/Amp(100μg/ml)液体培养基中使携带感兴趣质粒中的单集落生长过夜。用 1.0L 新鲜培养基以 1: 30 稀释细菌，让其在 30℃或 37℃生长直到 OD₅₅₀ 达到 0.6-0.8。用终浓度为 1.0mM 的 IPTG 诱导重组蛋白的表达。培育 3 小时后，在 4℃以 8000g 离心 15 分钟收获细菌，并重悬浮于 20ml 20mM Tris-HCl(pH 7.5)和完整的蛋白酶抑制剂(Boehringer-Mannheim)中。随后的所有步骤都是在 4℃
15 或冰上进行的。

用 Branson Sonifier 450 进行超声处理破碎细胞，并以 5000g 离心 20 分钟以沉积未破碎的细胞和包含体。以 50000g(Beckman Ti50, 29000rpm)离心包含膜和细胞碎片的上清液 75 分钟，用 20mM Bis-Tris 丙烷(pH 6.5)、1.0M NaCl、10%(v/v)甘油洗涤，并以 50000g 再次离心 75 分钟。将沉淀物重悬浮于 20mM
20 Tris-HCl(pH 7.5)、2.0%(v/v)十二烷基肌氨酸钠、完整的蛋白酶抑制剂(1.0mM EDTA 终浓度)，并培育 20 分钟以溶解内膜。以 5000g 离心沉淀的细胞碎片 10 分钟，以 75000g 离心上清液 75 分钟(Beckman Ti50, 33000rpm)。在上清液中发现蛋白质 008L 和 519L，表明内膜的定位。对这些蛋白质，用内膜和全膜组分(如上所述用 NaCl 洗涤)免疫接种小鼠。用 20mM Tris-HCl(pH 7.5)洗涤从 75000g
25 沉淀物得到的外膜小泡，并以 75000g 离心 75 分钟或过夜。最后将 OMV 重悬浮于 500μl 20mM Tris-HCl(pH 7.5)、10% v/v 甘油中。Orf1L 和 Orf40L 都定位于且在外膜组分中富含，将其用于免疫接种小鼠。用标准 Bradford 试验(Bio-Rad)计算蛋白质浓度，用 DC 蛋白质试验(Bio-Rad)确定内膜组分的蛋白质浓度。用 SDS-PAGE 分析分离步骤中的各种组分。

His-标记的蛋白质的纯化

从菌株 2996 和 MC58 中克隆了 287 的各种形式。用 C-末端 His 标记的融合体进行构建，其包括成熟形式(aa18-427)、含缺失($\Delta 1$ 、 $\Delta 2$ 、 $\Delta 3$ 和 $\Delta 4$)的构建物和由 B 或 C 结构域组成的克隆。对以 His-融合体纯化的各克隆而言，划线接种单集落，并且在 37°C LB/Amp(100 μ g/ml)琼脂板上培育过夜。将从该平板上分离的集落接种到 20ml LB/Amp(100 μ g/ml)的液体培养基中，并在 37°C 振荡生长过夜。以 1: 30 将过夜培养物稀释到 1.0L LB/Amp(100 μ g/ml)液体培养基中，让其在最佳温度(30 或 37°C)生长，直到 OD₅₅₀ 达到 0.6-0.8。添加 IPTG(终浓度为 1.0mM)诱导重组蛋白的表达，再培育培养物 3 小时。于 4°C，以 8000g 离心 15 分钟收获细菌。将细菌沉淀重物悬浮于 7.5ml(i)冷缓冲液 A(300mM NaCl、50mM 磷酸盐缓冲液、10mM 咪唑、pH 8.0)，用于可溶性蛋白质；或(ii)缓冲液 B(10mM Tris-HCl、100mM 磷酸盐缓冲液，pH 8.8 和任选地 8M 尿素)，用于不溶性蛋白质。以可溶形式纯化的蛋白质包括 287-His、 $\Delta 1$ 、 $\Delta 2$ 、 $\Delta 3$ 和 $\Delta 4$ 287-His、 $\Delta 4$ 287MC58-His、287c-His 和 287cMC58-His。蛋白质 287bMC58-His 是不溶的并相应地纯化。用 Branson Sonifier 450，在冰上以 40W、30 秒超声处理破坏细胞 4 次，并于 4°C 以 13000xg 离心 30 分钟。对于不溶蛋白质，将沉淀物重悬浮于 2.0ml 缓冲液 C(6M 盐酸胍、100mM 磷酸盐缓冲液、10mM Tris-HCl、pH 7.5)，并用 Dounce 匀浆器处理 10 次。以 13000g 离心匀浆 30 分钟并保留上清液。将可溶的和不可溶制备物的上清液与 150 μ l Ni²⁺-树脂(预先用缓冲液 A 或缓冲液 B 平衡)混合，并在室温温和振荡培育 30 分钟。树脂是按制造商说明制备的 Chelating Sepharose Fast Flow(Pharmacia)。于 4°C，以 700g 离心分批制备物 5 分钟，弃去上清液。用 10ml 缓冲液 A 或 B 洗涤树脂 2 次(分批)10 分钟，重悬浮于 1.0ml 缓冲液 A 或 B 中，加到一次性柱上。用(i)缓冲液 A(4°C)或(ii)缓冲液 B(室温)持续洗涤树脂，直到流出物的 OD₂₈₀ 达到 0.02-0.01。再用(i)冷缓冲液 C(300mM NaCl、50mM 磷酸盐缓冲液、20mM 咪唑、pH 8.0)或(ii)缓冲液 D(10mM Tris-HCl、100mM 磷酸盐缓冲液、pH 6.3 和任选地 8M 尿素)进一步洗涤树脂，直到流出物的 OD₂₈₀ 达到 0.02-0.01。加入 700 μ l(i)冷洗脱缓冲液 A(300mM NaCl、50mM 磷酸盐缓冲液、250mM 咪唑、pH 8.0)或(ii)洗脱缓冲

液 B(10mM Tris-HCl、100mM 磷酸盐缓冲液、pH 4.5 和任选地 8M 尿素)洗脱 His-融合蛋白, 收集组分直到 OD₂₈₀ 显示获得了所有重组蛋白。用 SDS-PAGE 分析 20μl 量的各洗脱组分。用 Bradford 试验计算蛋白质浓度。

5 变性的 His-融合蛋白的复性

需要变性以稳定 287bMC8, 因此在免疫接种前需要进行复性步骤。将甘油加到上述得到的变性组分中, 使终浓度为 10% v/v。用透析缓冲液 I(10% v/v 甘油, 0.5M 精氨酸、50mM 磷酸盐缓冲液、5.0mM 还原的谷胱甘肽、0.5mM 氧化的谷胱甘肽、2.0M 尿素, pH 8.8)将蛋白质稀释至 200μg/ml, 用相同的缓冲液在 4℃透析 12-14 小时。于 4℃, 用缓冲液 II(10% v/v 甘油, 0.5M 精氨酸、50mM 磷酸盐缓冲液、5.0mM 还原的谷胱甘肽、0.5mM 氧化的谷胱甘肽、pH 8.8)再进行透析 12-14 小时。用以下公式计算蛋白质的浓度:

$$\text{蛋白质(mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

15 氨基酸序列分析

根据制造商的推荐, 用配备在线乙内酰苯硫脲-氨基酸分析仪(System Gold)的 Beckman 测序仪(LF 3000)进行蛋白质 NH₂-末端的自动测序分析。

免疫接种

20 在第 0、21 和 35 天, 用抗原免疫接种 Balb/C 小鼠, 在第 49 天分析血清。

血清分析-ELISA

25 将不包裹的 MenB M7 和包裹的菌株置于巧克力琼脂板上, 在 37℃、5% CO₂ 中培育过夜。用无菌 Dracon 刷从琼脂板上收集细菌菌落, 并接种到含有 0.25% 葡萄糖的 Mueller-Hinton 肉汤(Difco)中。每 30 分钟监测一次细菌的生长, 随后测定 OD₆₂₀。让细菌生长直到 OD 达到 0.4-0.5。将培养物以 4000rpm 离心 10 分钟。弃去上清液, 用 PBS 洗涤细菌 2 次, 重悬浮于含有 0.025% 甲醛的 PBS 中, 并在 37℃培育 1 小时, 然后在 4℃搅拌培育过夜。在 96 孔 Greiner 板的各孔中添加 100μl 细菌细胞, 并在 4℃培育过夜。然后用 PBT 洗涤缓冲液(0.1%

- Tween-20 的 PBS 溶液)冲洗这些孔 3 次。在各孔中加入 200 μ l 饱和缓冲液(2.7 %聚乙烯吡咯烷酮 10 的水溶液), 将这些平板在 37 $^{\circ}$ C 培育 2 小时。用 PBT 冲洗这些孔 3 次。在各孔中加入 200 μ l 稀释的血清(稀释缓冲液: 1%BSA、0.1% Tween-20、0.1%NaN₃ 的 PBS 溶液), 将这些平板在 37 $^{\circ}$ C 培育 2 小时。用 PBT
- 5 冲洗这些孔 3 次。在各孔中加入 100 μ l HRP-缀合的兔抗-小鼠(Dako)血清(用稀释缓冲液以 1: 2000 稀释), 将这些平板置于 37 $^{\circ}$ C 培育 90 分钟。用 PBT 缓冲液洗涤这些孔 3 次。在各孔中加入 100 μ l HRP 的底物缓冲液(25ml 柠檬酸缓冲液 pH 5, 10mg 邻-苯二胺(phenildiamine)和 10 μ l H₂O₂), 并将平板在室温中静置 20 分钟。在各孔中加入 100 μ l 12.5% H₂SO₄, 随后测定 OD₄₉₀。计算 ELISA 滴定度,
- 10 即高于预先免疫血清稀释度的、OD490 值为 0.4 的血清稀释度。当 OD490 值为 0.4 的血清稀释度高于 1: 400 时, 将 ELISA 视为阳性。

血清分析-FACS 扫描细菌结合分析

- 将不包囊的 MenB M7 菌株置于巧克力琼脂板上, 在 37 $^{\circ}$ C、5%CO₂ 中培育
- 15 过夜。用无菌 Dracon 刷从琼脂板上收集菌落, 并接种到 4 支装有 0.25%葡萄糖的 8ml Mueller-Hinton 肉汤(Difco)试管中。每 30 分钟监测细菌的生长, 然后测定 OD₆₂₀。让细菌生长直到 OD 达到 0.35-0.5。将培养物以 4000rpm 离心 10 分钟。弃去上清液, 用封阻缓冲液(1%BSA 的 PBS 溶液, 0.4%NaN₃)重悬浮沉淀物, 并以 4000rpm 离心 5 分钟。将细胞重悬浮于封阻缓冲液, 使 OD₆₂₀ 达到 0.05。
- 20 在 96 孔 Costar 板的各孔中添加 100 μ l 细菌细胞。在各孔中加入 100 μ l 稀释的 (1:100、1: 200、1: 400)血清(以封阻缓冲液配制), 并将这些平板在 4 $^{\circ}$ C 培育 2 小时。以 4000rpm 离心细胞 5 分钟, 吸出上清液, 在各孔中加入 200 μ l/孔封阻缓冲液洗涤细胞。在各孔中加入 100 μ l R-Phicoerytrin 缀合的 F(ab)₂ 山羊抗-小鼠(以 1: 100 稀释), 将平板置于 4 $^{\circ}$ C 培育 1 小时。以 4000rpm 离心 5 分钟以沉
- 25 淀细胞, 加入 200 μ l/孔封阻缓冲液洗涤细胞。吸出上清液, 细胞重悬浮于 200 μ l/孔 PBS、0.25%甲醛。将样品转移到 FACScan 管中并读数。FACScan(Laser Power 15mW)的条件设为: FL2 开; FSC-H 阈值: 92; FSC PMT 电压: E01;SSC PMT:474;Amp. Gains 6.1;FL-2 PMT:586;补偿值: 0。

血清分析-杀菌试验

于 37℃、在 5%CO₂、在巧克力琼脂板(以冷冻原液起始)上培育脑膜炎奈瑟球菌 2996 过夜。收集菌落,并将其接种到 7ml 含有 0.25%葡萄糖的 Mueller-Hinton 肉汤中,使 OD₆₂₀ 达到 0.05-0.08。在 37℃振荡培育约 1.5 小时,直到 OD₆₂₀ 达到 0.23-0.24。用 50mM 磷酸盐缓冲液(pH7.2, 含 10mM MgCl₂、10mM CaCl₂ 和 0.5%(w/v)BSA(分析缓冲液))以 10⁵CFU/ml 工作稀释度稀释细菌。最终反应混合物的总体积为 50μl,其中 25μl 连续 2 倍稀释的测试血清,12.5μl 工作稀释度的细菌,12.5μl 幼兔补体(终浓度 25%)。

对照包括:用补体血清培育的细菌、用细菌培育并在 56℃加热 30 分钟补充灭活的免疫血清。在加入幼兔补体后,用斜置方法立即将 10μl 对照置于 Mueller-Hinton 琼脂板上(0 时间)。37℃旋转培育 96 孔板 1 小时。将每份样品的 7μl 涂布在 Mueller-Hinton 琼脂板上作为斑点,而用斜置方法将 10μl 对照涂布在 Mueller-Hinton 琼脂板上(1 时间)。37℃培育琼脂板 18 小时,计算相应于 0 时间和 1 时间的菌落数量。

15

血清分析-Western 印迹法

将纯化的蛋白质(500ng/泳道)、外膜小泡(5μg)和 MenB 菌株 2996 衍生的全细胞提取物(25μg)加到 12%SDS-聚丙烯酰胺凝胶上,并转移到硝基纤维素膜上。用转化缓冲液(0.3% Tris 碱, 1.14%甘氨酸、20%(v/v)甲醇)在 4℃、150mA 进行转化 2 小时。于 4℃在饱和缓冲液(10%脱脂乳、0.1%Triton X100 的 PBS 溶液)中培育过夜使膜饱和。用冲洗缓冲液(3%脱脂乳、0.1%Triton X100 的 PBS 溶液)冲洗膜 2 次,并在 37℃与用洗涤缓冲液以 1:200 稀释的小鼠血清一起培育 2 小时。冲洗膜 2 次,与 1:2000 稀释的辣根过氧化物酶标记的抗-小鼠 Ig 一起培育 90 分钟。用 0.1% Triton X100 的 PBS 溶液冲洗膜 2 次,用 Opti-4CN 底物试剂盒(Bio-Rad)显色。加入水终止反应。

25

如下制备 OMV: 37℃、5%CO₂, 在 5 个 GC 板上让脑膜炎奈瑟球菌 2996 生长过夜,用接种环收获,并重悬浮于 10ml 20mM Tris-HCl pH 7.5、2mM EDTA 中。在 56℃热灭活 45 分钟,在冰上超声处理破碎细胞 5 分钟(50%负载循环(duty cycle)、50%输出, Branson 超声仪 3mm 微型针头)。以 5000g 离心 10 分钟除去

未破碎的细胞，回收含有完整细胞包膜组分的上清液，于 4℃ 以 50000g 进一步离心过夜。将含有膜的沉淀物重悬浮于 2% 二烷基肌氨酸、20mM Tris-HCl pH 7.5、2mM EDTA 中，在室温培育 20 分钟以溶解内膜。以 10000g 离心上清液 10 分钟除去聚集体，以 50000g 进一步离心上清液 3 小时。用 PBS 冲洗含有外膜的沉积物，并重悬浮于相同的缓冲液中。用 BSA 作为标准，由 D.C.Bio-Rad 蛋白分析(改进的 Lowry 方法)测定蛋白质的浓度。

如下制备全细胞提取物：使脑膜炎奈瑟球菌在 GC 板上培育过夜，用接种环收获，并重悬浮于 1ml 20mM Tris-HCl 中。在 56℃ 热灭活 30 分钟。

10 961 结构域的研究

细胞组分的制备 用过夜培养物或在 IPTG 培育 3 小时后的细菌制备表达 961 不同结构域的大肠杆菌克隆的全裂解物、周质、上清液和 OMV。简单地说，获得周质，细菌悬浮于蔗糖 25% 和 Tris 50mM(pH 8)(含有 polymyxine 100μg/ml)。1 小时后，室温以 13000rpm 离心细菌 15 分钟，收集上清液。用 0.2μm 过滤培养物上清液，在冰中用 TCA 50% 沉淀 2 小时。离心(13000rp, 30 分钟)后，用乙醇 70% 漂洗沉淀物 2 次，并悬浮于 PBS 中。如上所述，进行 OMV 制备。用针对 GST-961 的多克隆抗血清，在 SDS-PAGE 或 Western 印迹中分析各细胞组分。

20 粘附分析 在补充有 10% 热灭活的 PCS、15mM L-谷氨酰胺和抗生素的 DMEM(Gibco)中维持张氏 (Chang) 上皮细胞(Wong-Kilbourne 衍生物，克隆 1-5c-4，人结膜)。

为了进行粘附分析，用 PBS 漂洗张氏上皮细胞的亚融合培养物，并用胰蛋白酶-EDTA(Gibco)处理，从而从塑性支持体上释放它们。然后将细胞悬浮于 PBS 中，计数并在 PBS 中稀释至 5×10^5 细胞/ml。

沉淀过夜培育的或用 IPTG 诱导后的细菌，并用 PBS 通过以 13000 离心 5 分钟洗涤 2 次。在室温、暗处，在 1ml 缓冲液(含有 50mM NaHCO₃ 和 100mM NaCl pH 8)中将约 $2-3 \times 10^8$ (cfu)与 0.5mg/ml FITC(Sigma)一起培育 30 分钟。洗涤 FITC 标记的细菌 2-3 次，并以 $1-1.5 \times 10^9$ /ml 悬浮于 PBS 中。于 37℃，将 200μl 此悬

浮液($2-3 \times 10^8$)与 $200\mu\text{l}$ (1×10^5)上皮细胞一起培育 30 分钟。然后以 2000rpm 离心细胞 5 分钟，除去未粘附的细胞，悬浮液 $200\mu\text{l}$ PBS 中，转移到 FACScan 管中并读数。

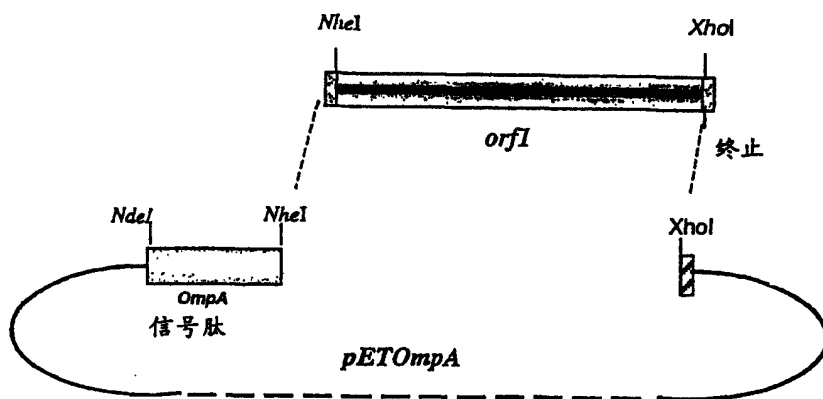


图 1

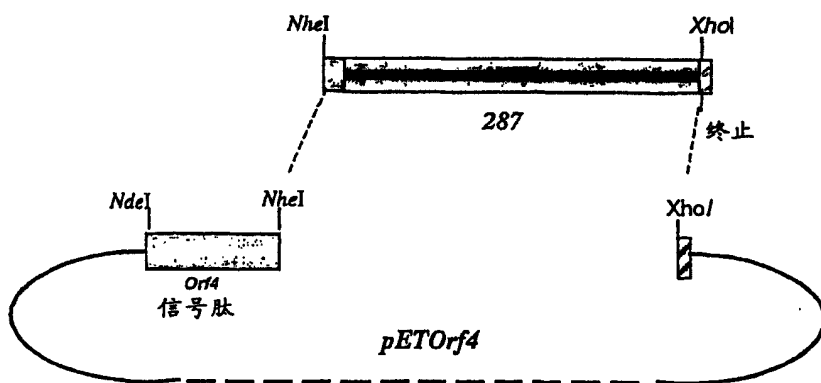


图 2

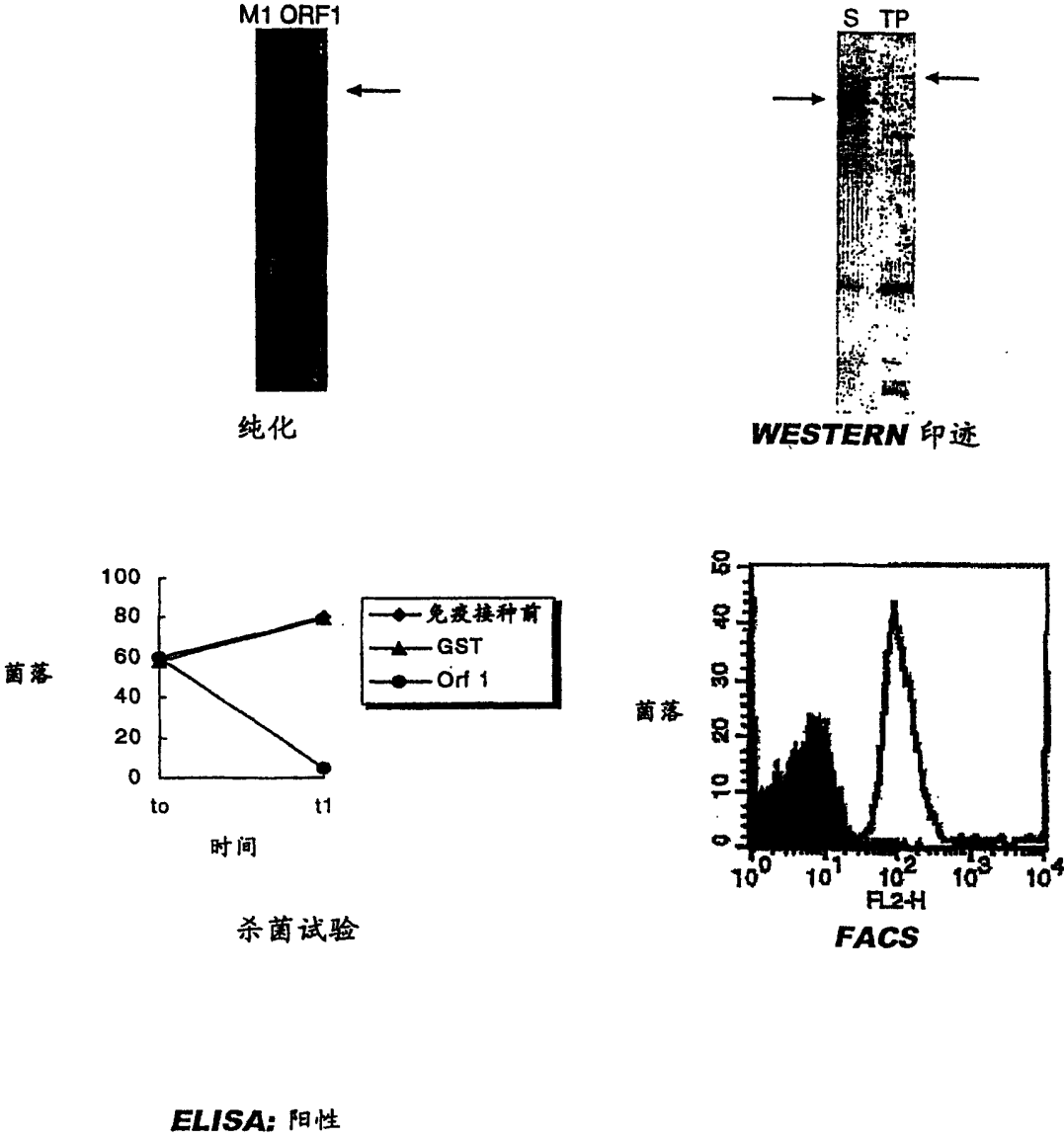
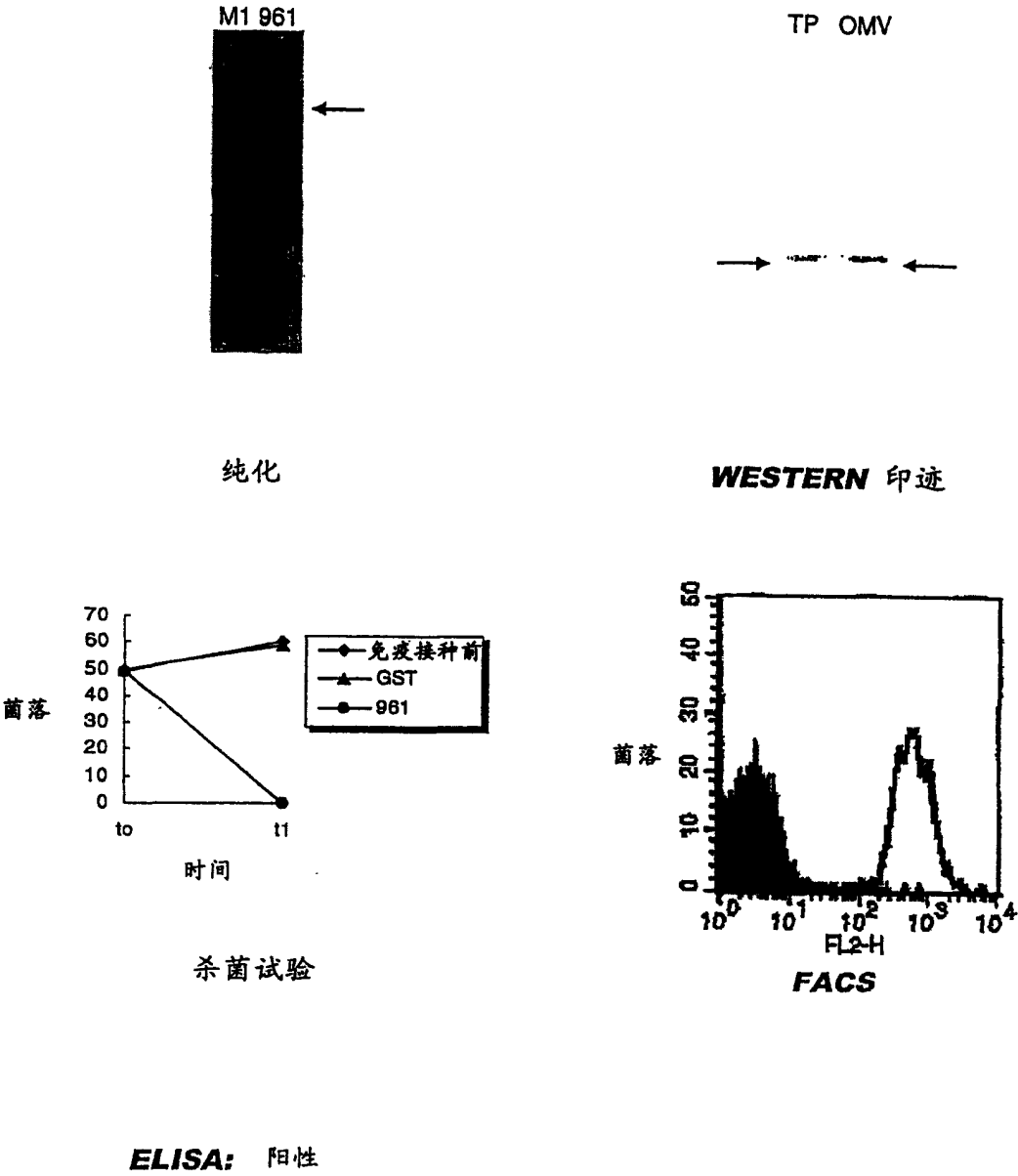


图 3



图

4

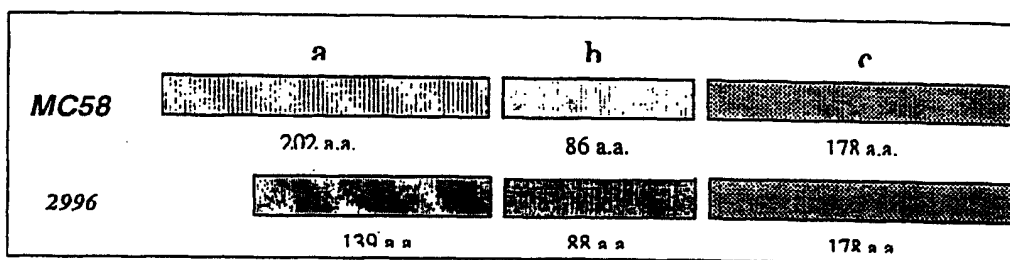


图 5

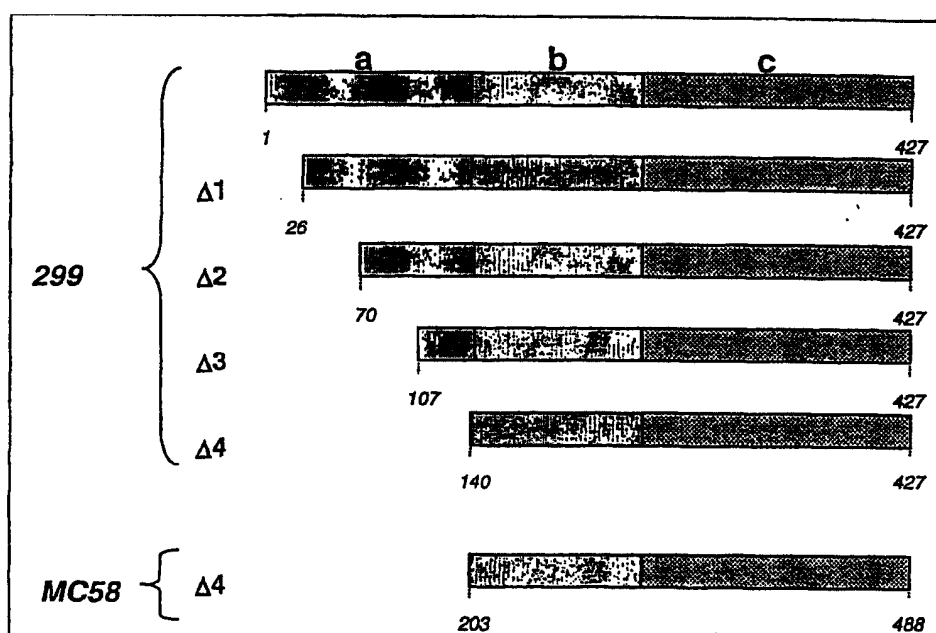


图 6

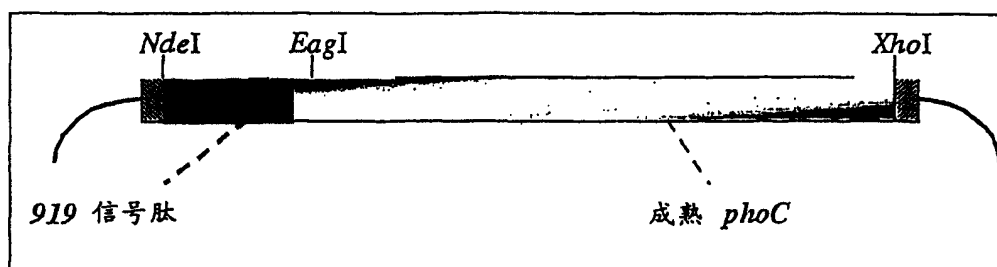


图 9

<A-----<A1
 MC58 1 MIER SVIAMACIFALSACGGGGGSPDVKSADTLSPAPVWSEKETPAKEDAPQAGSQG
 2996 1 MIER SVIAMACIFALSACGGGGGSPDVKSADTLSPAPVWSEKETPAKEDAPQAGSQG

-----<A2
 MC58 61 QCAPSAQGSQDMAVSEENTGNGCAVTADNPKNEDEVAQNDHPQNAAGTDSSTPNHTPDF
 2996 61 QCAPSTQGSQDMAVSAENTGNGCAVTDPKPKNEDEGPQNDHPQ.....

-----<A3
 MC58 121 NMLAGNMENQATDAGESSQPANQPDMANAADGMQGGDPSAGGQNAAGNTAAQGANQAGNNQ
 2996 106 SAESANOTGNAG

-----A><B-----
 MC58 181 AAGSSDPIPASNPAPANGGSNEGRVDLANGLIDGPSONITLTHCGDSCSNFDEEV
 2996 118 PADSSDSAPASNPAPANGGSNEGRVDLANGLIDGPSONITLTHCGDSCSNGLDEEA

-----B>
 MC58 241 QLKSEFEKLSADAKLSTYKNDCKNDKEVGLVADSVQMKGINOYLIFYPK..PTSTARFR
 2996 178 PSKSEFEKLSNESERIEKTKNDCKSDKEINLVATAVOANGINKYVLIYKDKSASSSARFR

-----<C-----
 MC58 299 RSARSRRSLPAEMPLIPVQADTLVDGEAVSLTCHSCNIFAPEGNTYRLTVGAENLPGG
 2996 238 RSARSRRSLPAEMPLIPVQADTLVDGEAVSLTCHSCNIFAPEGNTYRLTVGAENLPGG

 MC58 359 SYALRVQGEPAKCEMLAGTAVTNGEVLHFTHTENGHPYPTGRFAAKWDFGSKSVDCIIDS
 2996 298 SYALRVQGEPAKCEMLAGTAVTNGEVLHFTHTENGHPYPTGRFAAKWDFGSKSVDCIIDS

 MC58 419 GDDLHMGTOKEKAAIDGNGFRGTNTENGSGDVSCRFYGFAGEEVAGKYSYRPTDAEKGGF
 2996 358 GDDLHMGTOKEKAAIDGNGFRGTNTENGSGDVSCRFYGFAGEEVAGKYSYRPTDAEKGGF

-----C>
 MC58 479 GVFA GKKEQD*
 2996 418 GVFA GKKEQD*

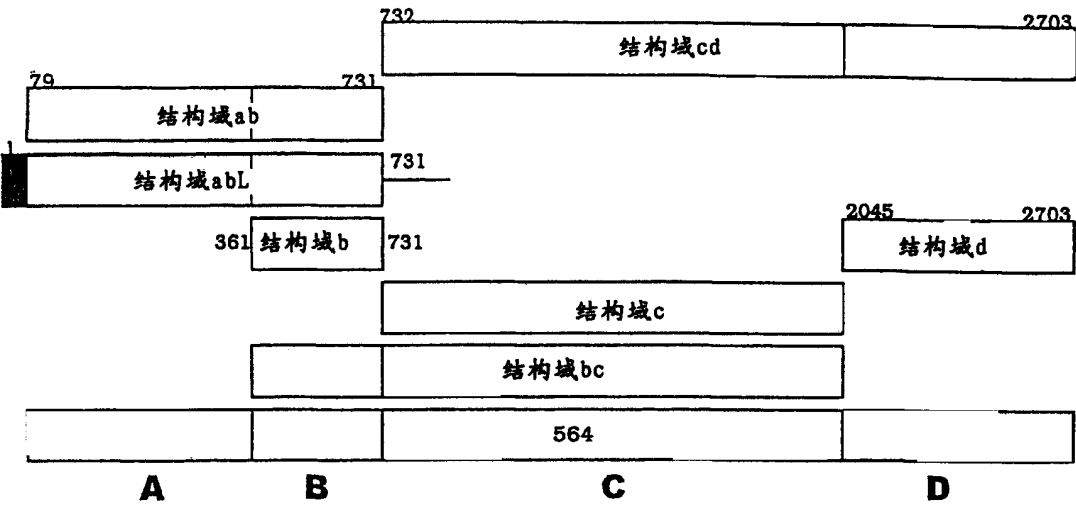


图 8

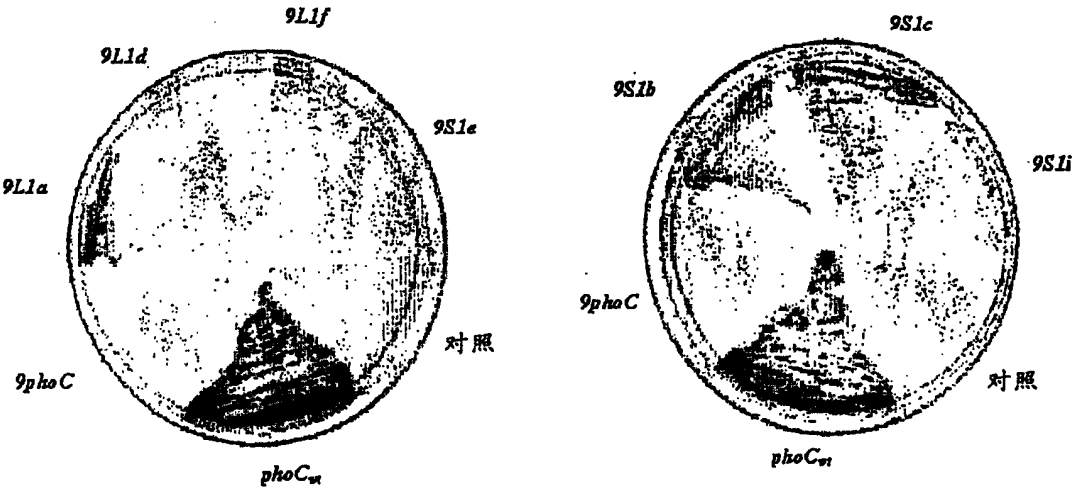


图 10

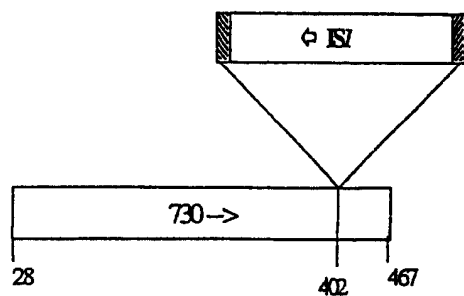


图 11A

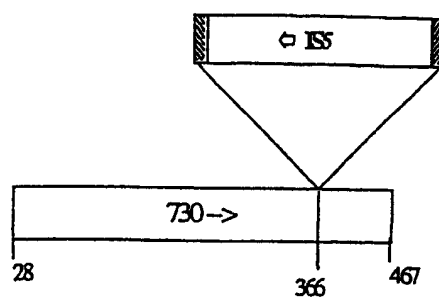
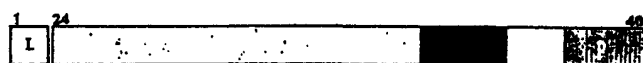


图 11B

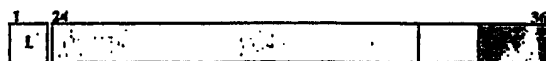
961 (2996)

961 L (2996) □

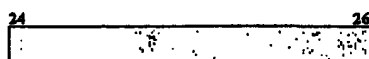


961 (MC58)

961 L (MC58) □



961a (2996=MC58)

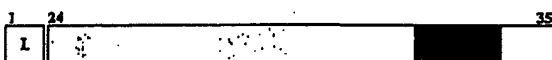


961b (2996)



961c (2996)

961c-L (2996) □

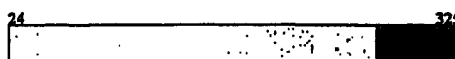


961c (MC58)

961c-L (MC58) □

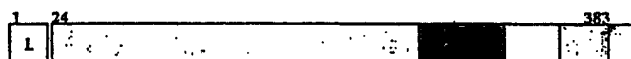


961d (2996)



961-Δ1 (2996)

961Δ1-L □



□ 前导肽

■ 2996中的, 而非MC58中的区域

□ 卷曲螺旋的片段

■ 膜锚

图 12

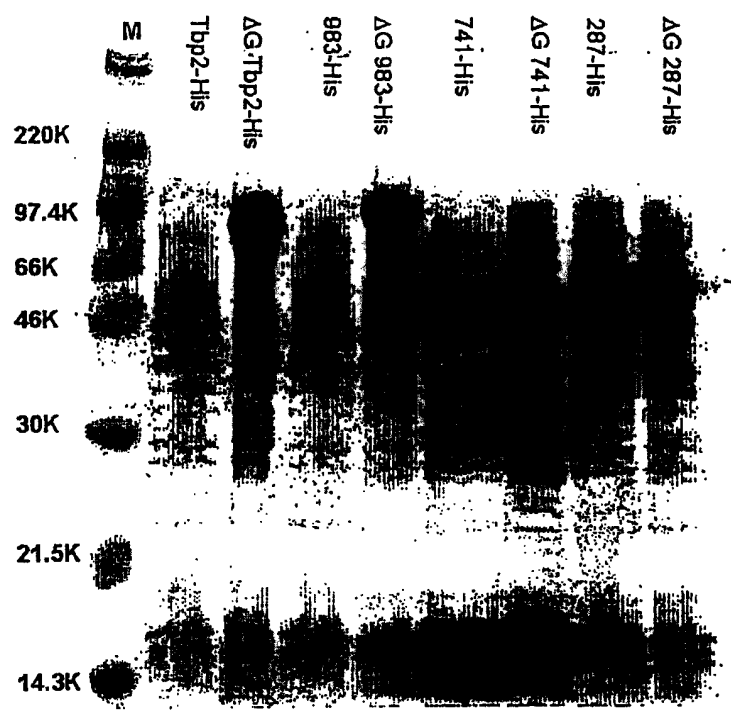


图 13

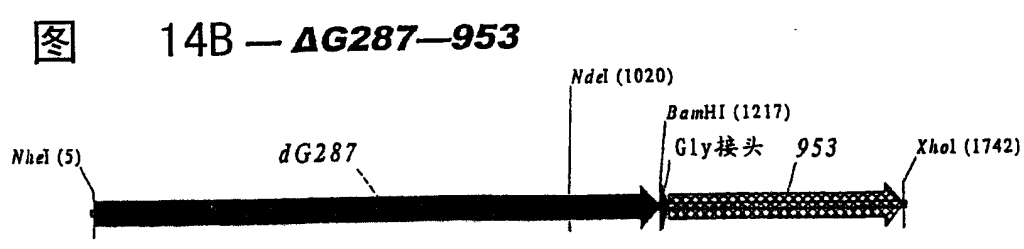
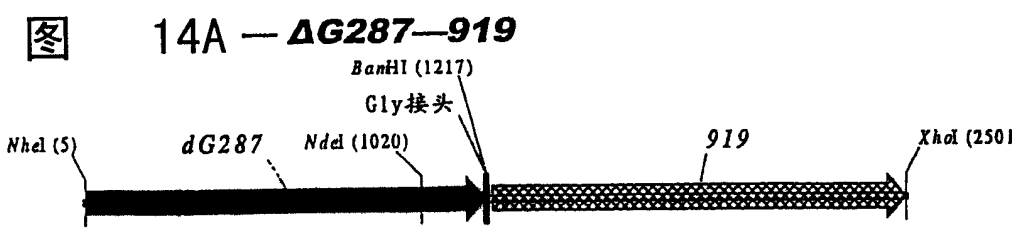


图 14C — $\Delta G287-961$



图 14D — $\Delta G287NZ-919$

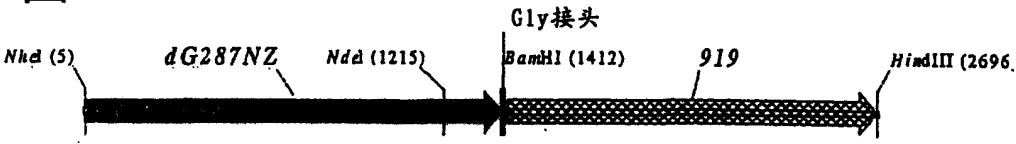


图 14E — $\Delta G287NZ-953$



图 14F — $\Delta G287NZ-961$

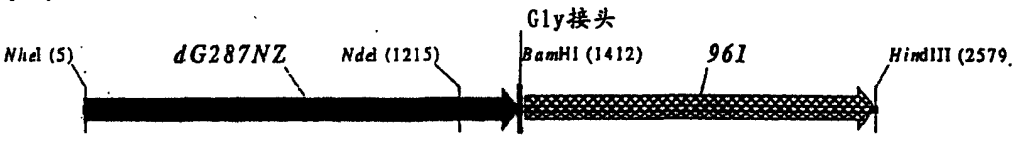


图 14G — $\Delta G983-ORF46.1$



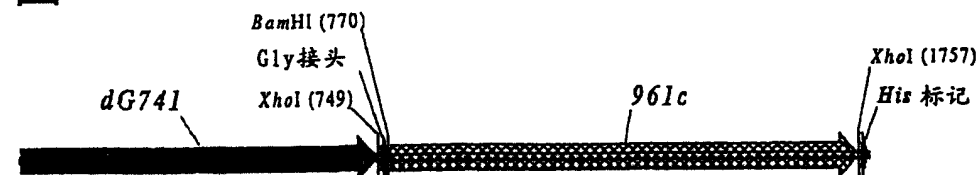
图 14H — $\Delta G983-741$ 图 14I — $\Delta G983-961$ 图 14J — $\Delta G983-961c$ 图 14K — $\Delta G741-961$ 图 14L — $\Delta G741-961c$ 

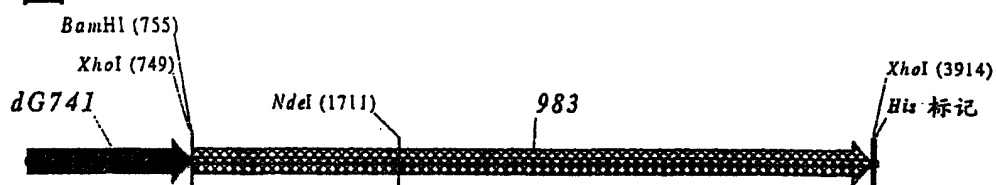
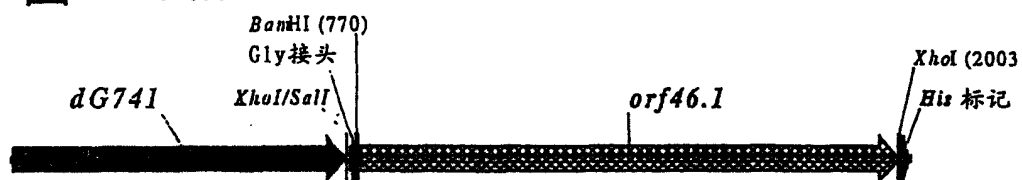
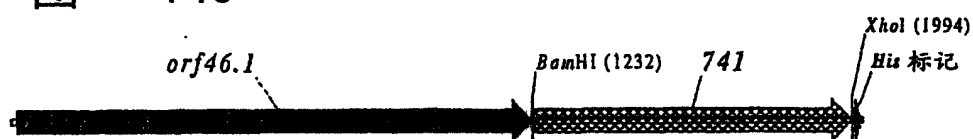
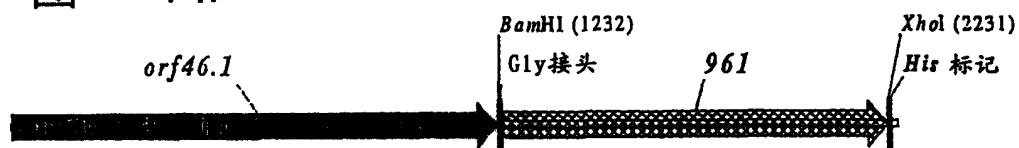
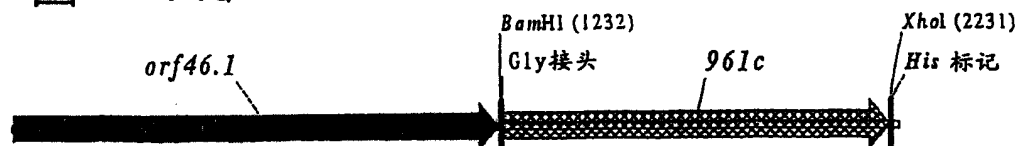
图 14M — $\Delta G741-983$ 图 14N — $\Delta G741-ORF46.1$ 图 14O — $ORF46.1-741$ 图 14P — $ORF46.1-961$ 图 14Q — $ORF46.1-961c$ 

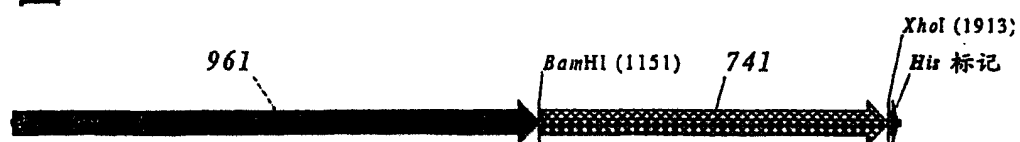
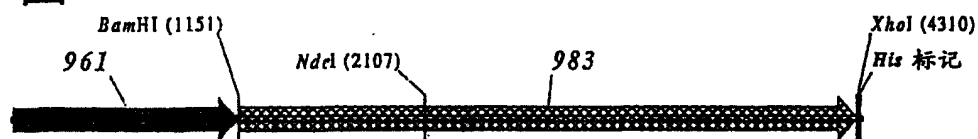
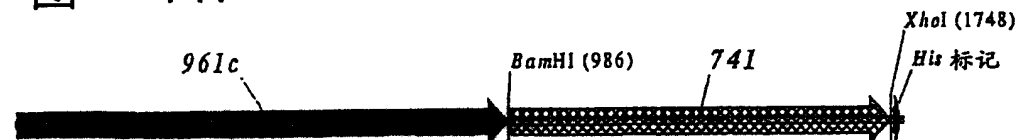
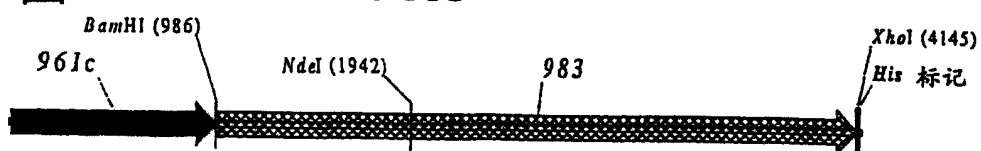
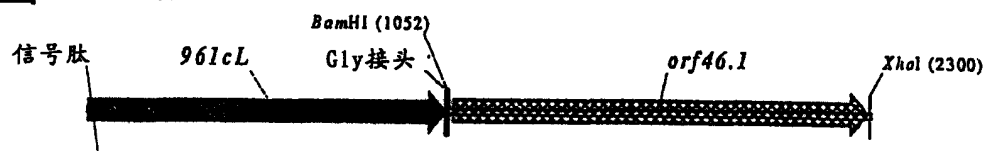
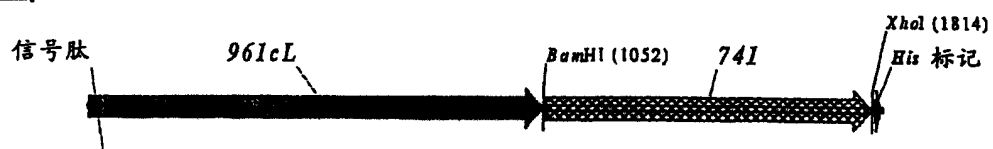
图 14R — 961-ORF46.1**图 14S — 961-741****图 14T — 961-983****图 14U — 961c-ORF46.1****图 14V — 961c-741**

图 14W — 961c-983**图 14X — 961cL-ORF46.1****图 14Y — 961cL-741****图 14Z — 961cL-983**