



(12)发明专利

(10)授权公告号 CN 102724997 B

(45)授权公告日 2016.09.21

(21)申请号 201080061239.X

(22)申请日 2010.11.16

(65)同一申请的已公布的文献号
申请公布号 CN 102724997 A

(43)申请公布日 2012.10.10

(30)优先权数据
61/272,931 2009.11.19 US

(85)PCT国际申请进入国家阶段日
2012.07.12

(86)PCT国际申请的申请数据
PCT/US2010/002980 2010.11.16

(87)PCT国际申请的公布数据
W02011/062615 EN 2011.05.26

(73)专利权人 格林考瓦因有限公司
地址 瑞士施利伦

(72)发明人 M·瓦克 查尔斯·韦克特

(74)专利代理机构 中国专利代理(香港)有限公司
72001

代理人 梁谋 彭昶

(51)Int.Cl.

A61K 39/385(2006.01)

A61K 39/02(2006.01)

A61K 39/108(2006.01)

C07K 1/00(2006.01)

(56)对比文件

WO 2009104074 A2,2009.08.27,

WO 2009104074 A2,2009.08.27,

DMITRIEV等.Somatic antigens of Shigella. The structure of the specific polysaccharide of Shigella newcastle (Sh. flexneri type 6) lipopolysaccharide.《EUR. J. BIOCHEM》.1979,(第98期),

RUSH等.A novel epimerase that converts GlcNAc-P-P-undecaprenol to GalNAc-P-P- undecaprenol in Escherichia coli 0157.《J. BIOL. CHEM》.2009,第285卷(第3期),

审查员 陈云华

权利要求书2页 说明书76页

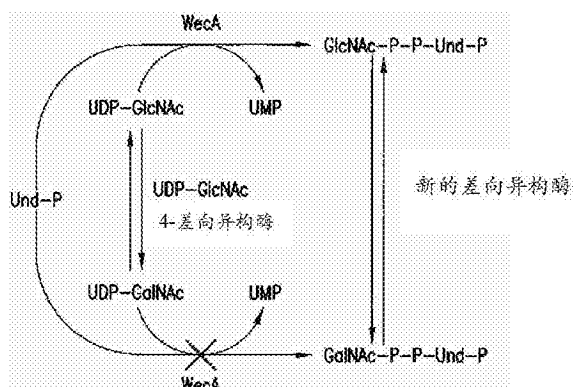
序列表51页 附图12页

(54)发明名称

在原核细胞中产生免疫原性多糖的生物合成系统

(57)摘要

本发明涉及包含N-糖基化蛋白的生物缀合物疫苗。此外,本发明涉及一种重组原核生物合成系统,包含编码差向异构酶的核酸,该差向异构酶合成在还原末端具有N-乙酰基半乳糖胺的寡糖或多糖。本发明还涉及含有在还原末端具有N-乙酰基半乳糖胺的寡糖或多糖的N-糖基化蛋白及用于产生这类N-糖基化蛋白的表达系统和方法。



1. 一种重组原核生物合成系统,其包含编码以下酶和蛋白的核酸:

在十一异戊二烯基焦磷酸上合成N-乙酰基半乳糖胺的差向异构酶,其中编码所述差向异构酶的核酸选自SEQ ID NO.1、SEQ ID NO.3、SEQ ID NO.5、SEQ ID NO.7和SEQ ID NO.9;

在脂质载体上装配寡糖或多糖的糖基转移酶;

寡糖基转移酶;和

包含引入的共有序列D/E-X-N-Z-S/T中的一个或多个的蛋白,其中X和Z独立地为除脯氨酸以外的任何天然氨基酸。

2. 权利要求1的重组原核生物合成系统,其中所述寡糖基转移酶是来自空肠弯曲杆菌(*Campylobacter jejuni*)。

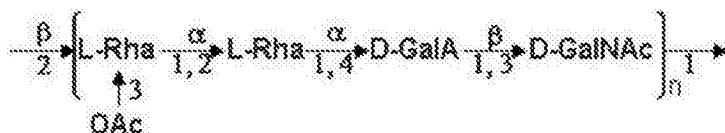
3. 权利要求1的重组原核生物合成系统,其中所述蛋白是铜绿假单胞菌胞外蛋白。

4. 权利要求1的重组原核生物合成系统,其中所述寡糖或多糖是来自革兰氏阴性菌。

5. 权利要求4的重组原核生物合成系统,其中所述寡糖或多糖是来自弗氏志贺氏菌(*Shigella flexneri*)。

6. 权利要求5的重组原核生物合成系统,其中所述寡糖或多糖是来自弗氏志贺氏菌6。

7. 权利要求1的重组原核生物合成系统,其中所述寡糖或多糖包含以下结构:



8. 权利要求4的重组原核生物合成系统,其中所述寡糖或多糖是来自大肠杆菌。

9. 权利要求8的重组原核生物合成系统,其中所述寡糖或多糖是来自大肠杆菌0157。

10. 权利要求1的重组原核生物合成系统,其中所述寡糖或多糖包含结构 α -D-PerNAc- α -L-Fuc- β -D-Glc- α -D-GalNAc。

11. 一种产生N-联糖基化蛋白的方法,包括:

a.) 将编码i.)~iv.)的核酸导入宿主生物中:

i.) 在十一异戊二烯基焦磷酸上合成N-乙酰基半乳糖胺的差向异构酶,其中编码所述差向异构酶的核酸选自SEQ ID NO.1、SEQ ID NO.3、SEQ ID NO.5、SEQ ID NO.7和SEQ ID NO.9;

ii.) 在脂质载体上装配寡糖或多糖的糖基转移酶;

iii.) 寡糖基转移酶;和

iv.) 包含引入的共有序列D/E-X-N-Z-S/T中的一个或多个的蛋白质,其中X和Z独立地为除脯氨酸以外的任何天然氨基酸;和

b.) 培养所述宿主生物直到至少一种N-糖基化蛋白产生。

12. 权利要求11的方法,其中所述寡糖基转移酶是来自空肠弯曲杆菌。

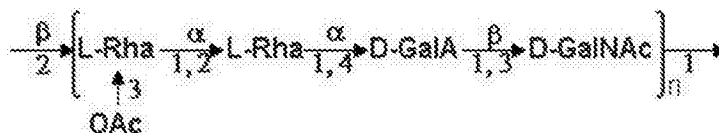
13. 权利要求11的方法,其中所述蛋白是铜绿假单胞菌胞外蛋白。

14. 权利要求11的方法,其中所述寡糖或多糖是来自革兰氏阴性菌。

15. 权利要求14的方法,其中所述寡糖或多糖是来自弗氏志贺氏菌。

16. 权利要求15的方法,其中所述寡糖或多糖是来自弗氏志贺氏菌6。

17. 权利要求11的方法,其中所述寡糖或多糖包含以下结构:



18. 权利要求14的方法,其中所述寡糖或多糖是来自大肠杆菌。

19. 权利要求18的方法,其中所述寡糖或多糖是来自大肠杆菌0157。

20. 权利要求11的方法,其中所述寡糖或多糖包含结构 α -D-PerNAc- α -L-Fuc- β -D-Glc- α -D-GalNAc。

21. 权利要求11的方法,进一步包括从培养物分离所述N-糖基化蛋白。

22. 一种重组原核生物合成系统,其包含编码以下酶和蛋白的核酸:

在十一异戊二烯基焦磷酸上合成N-乙酰基半乳糖胺的差向异构酶,其中编码所述差向异构酶的核酸是SEQ ID NO.1;

在脂质载体上装配寡糖或多糖的糖基转移酶;

寡糖基转移酶;和

包含引入的共有序列D/E-X-N-Z-S/T中的一个或多个的蛋白,其中X和Z独立地为除脯氨酸以外的任何天然氨基酸。

23. 一种产生N-联糖基化蛋白的方法,包括:

a.) 将编码i.)~iv.)的核酸导入宿主生物中:

i.) 在十一异戊二烯基焦磷酸上合成N-乙酰基半乳糖胺的差向异构酶,其中编码所述差向异构酶的核酸是SEQ ID NO.1;

ii.) 在脂质载体上装配寡糖或多糖的糖基转移酶;

iii.) 寡糖基转移酶;和

iv.) 包含引入的共有序列D/E-X-N-Z-S/T中的一个或多个的蛋白质,其中X和Z独立地为除脯氨酸以外的任何天然氨基酸;和

b.) 培养所述宿主生物直到至少一种N-糖基化蛋白产生。

在原核细胞中产生免疫原性多糖的生物合成系统

[0001] 相关申请的交叉引用

[0002] 根据35 U.S.C. §119(e), 本申请要求于2009年11月19日提交的美国临时专利申请第61/272,931号的权益, 该美国临时专利申请通过引用其全部结合到本文中。

发明领域

[0003] 本发明涉及生物合成系统和蛋白质在制备疫苗中的应用。另外, 本发明涉及重组原核生物合成系统, 该系统具有启动在还原末端具有特定单糖的寡糖或多糖(oligo- or polysaccharide)的合成的差向异构酶。本发明还涉及在表达系统中用聚糖产生的N-糖基化蛋白及由包含免疫原性聚糖的所述N-糖基化蛋白制成的生物缀合物疫苗(bioconjugate vaccines), 并且提供用于产生N-糖基化蛋白的方法。

[0004] 发明背景

[0005] 糖蛋白是具有一个或多个共价连接的糖聚合体的蛋白质。N-联蛋白质糖基化是在真核生物的内质网中发生的一种必需和保守的过程。它对于分泌蛋白和膜蛋白的蛋白质折叠、寡聚化、稳定性、质量控制、分选(sorting)和转运来说是十分重要的(Helenius, A. 和 Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum (N-联聚糖在内质网中的作用). Annu. Rev. Biochem. 73, 1019-1049)。

[0006] 蛋白质糖基化对于蛋白质的免疫原性、稳定性和半寿期具有深远的影响。另外, 糖基化可有助于蛋白质的纯化, 即利用与固相结合并与蛋白质的糖基化部分相互作用的凝集素配体进行层析(例如亲和层析)来纯化蛋白质。因此, 为了提供生物学和药学上有用的糖基化模式, 惯例是在真核细胞中重组产生许多糖基化蛋白。

[0007] WO 2003/07467 (Aebi 等人) 证明了经食物传播的病原体空肠弯曲杆菌(*Campylobacter jejuni*) (它是一种细菌) 可以使其蛋白质N-糖基化, 除了古生菌(archaea)的某些种以外, 在已知的原核生物中, 这是一种独特的特征。糖基化所需要的机器由12个基因编码, 这12个基因在所谓的pgl基因座中是成簇的。N-糖基化的破坏会影响空肠弯曲杆菌(*C. jejuni*)的侵袭和致病机制但并不是致死的, 正如在大多数真核生物中(Burda P. 和 M. Aebi, (1999). The dolichol pathway of N-linked glycosylation (N-联糖基化的多萜醇途径). Biochem Biophys Acta 1426(2): 239-57)。在大肠杆菌(*E. coli*)中通过同时重组表达pgl基因座和受体糖蛋白来重建空肠弯曲杆菌蛋白的N-糖基化是可能的(Wacker 等人(2002). N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli* (空肠弯曲杆菌中的N-联糖基化及其功能性转移到大肠杆菌中). Science 298, 1790-1793)。

[0008] N-聚糖具有一种与蛋白质中的共有序列连接的聚糖。蛋白质中已知的N-糖基化共有序列允许原核生物中重组靶蛋白的N-糖基化。这类生物体包含寡糖基转移酶(“OT”; “OTase”), 例如空肠弯曲杆菌的寡糖基转移酶, 寡糖基转移酶是一种将聚糖转移到蛋白质的共有序列上的酶。

[0009] WO 2003/07467 (Aebi 等人) 教导了一种被导入了编码以下(i)、(ii)和(iii)的核

酸的原核生物:(i)在脂质载体上装配寡糖的特定糖基转移酶,(ii)包含共有序列“N-X-S/T”的重组靶蛋白,其中X可以为除脯氨酸以外的任何氨基酸,和(iii)寡糖基转移酶,例如将所述寡糖与靶蛋白的共有序列共价连接的空肠弯曲杆菌寡糖基转移酶。所述原核生物产生具有由特定糖基转移酶类型限定的特定结构的N-聚糖。

[0010] WO 2006/119987(Aebi等人)描述了用于在原核生物体内(in vivo)高效N-糖基化的蛋白质,以及高效产生N-糖基化蛋白质的手段(means)和方法。它进一步描述了将N-聚糖有效引入到重组蛋白中以改变所述蛋白的免疫原性、稳定性、生物活性、预防和/或治疗活性,并且提供了在其表面有效地展示本发明的重组N-糖基化蛋白的宿主细胞。另外,它描述了一种重组N-糖基化蛋白,其包含以下N-糖基化最优化氨基酸序列中的一个或多个:

[0011] D/E-X-N-Z-S/T(最优化共有序列),

[0012] 其中X和Z可以为除Pro以外的任何天然氨基酸,并且其中所述N-糖基化部分氨基酸序列中的至少一个被引入。将特定部分氨基酸序列(最优化共有序列)引入到蛋白质会导致产生在这些引入的位置中被寡糖基转移酶有效地N-糖基化的蛋白质。

[0013] 不同多糖的生物合成在细菌细胞中是保守的。多糖在载体脂质(carrier lipid)上从细胞质膜上的共同前体(活化糖核苷酸)由具有确定特异性的不同糖基转移酶进行装配。脂多糖(“LPS”)仅在革兰氏阴性菌例如志贺氏菌(*Shigella* spp.)、假单胞菌(*Pseudomonas* spp.)和大肠杆菌(*E. coli*)(ExPEC,EHEC)中提供。

[0014] LPS的合成始于在膜的细胞质一侧将单糖添加到载体脂质十一异戊二烯基磷酸(undecaprenyl phosphate,“Und-P-P”)上。通过不同糖基转移酶序贯添加来自活化糖核苷酸的单糖构建抗原,而脂质-连接的多糖通过翻转酶(flippase)穿过膜转向。抗原-重复单位通过酶反应进行聚合。然后多糖通过连接酶(Ligase)WaaL转移到脂质A,形成LPS,再输出到表面,而荚膜多糖在聚合后从载体脂质中释放出来并输出到表面。这些多糖的生物合成途径使得LPS生物缀合物能够在体内(in vivo)产生,俘获周质中的多糖到蛋白载体。

[0015] 寡糖或多糖(即糖残基)和蛋白(即蛋白载体)的这类合成复合物可以用作缀合物疫苗以抵抗多种细菌感染。缀合物疫苗已经成功地用于抵抗细菌感染。抗原性多糖与蛋白载体的缀合需要保护性记忆应答,因为多糖是不依赖T-细胞的免疫原。利用多糖以及蛋白载体中的活化反应基团,通过不同的化学方法已经使多糖与蛋白载体缀合。

[0016] 缀合物疫苗可以给予儿童以抵抗细菌感染并且也可以给成年人提供长时间持续的免疫应答。WO 2009/104074(Fernandez等人)的构建体已经发现在动物中产生IgG应答。在人类中,已经发现针对志贺氏菌属O-特异性多糖-蛋白缀合物疫苗的IgG应答与人体内的免疫保护有关(Passwell,J.H.等人,“Safety and Immunogenicity of Improved *Shigella* O-Specific Polysaccharide-Protein Conjugate Vaccines in Adults in Israel(改进的志贺氏菌属O-特异性多糖-蛋白缀合物疫苗在以色列成年人中的安全性和免疫原性)”*Infection and Immunity*,69(3):1351-1357(2001年3月))。据信多糖(即糖残基)触发糖特异性的短期免疫应答。事实上,人体的免疫系统对细菌的特定多糖表面结构例如O-抗原和荚膜多糖会产生强烈应答。然而,由于对多糖的免疫应答是依赖IgM的,因此免疫系统不产生记忆。携带多糖的蛋白载体能触发IgG应答,IgG应答是依赖T-细胞的并提供长时间持续的保护作用,因为免疫系统产生了记忆。

[0017] 大肠杆菌0157是一种引起溶血性尿毒综合征所有当前病例三分之二左右的肠出

血性菌株并造成严重的人类健康担忧(Law,D.(2000)J.App.Microbiol.,88,729-745; Wang,L.和Reeves,P.R.(1998)Infect.Immun.66,3545-3551)。

[0018] 大肠杆菌菌株0157(*Escherichia coli* strain 0157)产生含有重复四糖单元(4-N-乙酰基 perosamine→岩藻糖→葡萄糖→GalNAc)(α -D-PerNAc- α -L-Fuc- β -D-Glc- α -D-GalNAc)的O-抗原(Perry,M.B.,MacLean,L.和Griffith,D.W.(1986)Biochem.Cell.Biol., 64,21-28)。该四糖在十一异戊二烯基焦磷酸上预装配。大肠杆菌细胞被膜含有内浆膜即带有应力的(stress-bearing)肽聚糖层和由磷质内单层和外单层组成的不对称外膜,这构成了细菌LPS。LPS含有脂质A锚形体、含有3-脱氧-D-甘露-辛-2-酮糖酸的核心和O-抗原区三个组分(参见:Raetz,C.R.H.和Whitfield,C.(2002)Annu.Rev.Biochem.,71,635-700; Whitfield,C.(2006)Ann.Rev.Biochem.75,39-68;Samuel,G.和Reeves,P.R.(2003)Carbohydrate Research,338,2503-2519;及其中关于细菌LPS的O-抗原装配综述的参考文献)。

[0019] 细菌LPS的O-抗原组分是大的、极度多变的多糖,可以是同聚的,由单一重复单糖组成,或者杂聚的,含有3-6个糖单元的10-30个重复(Reeves,P.R.,Hobbs,M.,Valvano,M.A.,Skurnik,M.,Whitfield,C.,Coplin,D.,Kido,N.,Klena,J.,Maskell,D.,Raetz,C.R.H.和Rick,P.D.(1996)Trends Microbiol.,4,495-503)。因此,O-抗原是细菌细胞表面的主要特征并构成毒力和致病性的重要决定因子(Law,D.(2000)J.App.Microbiol.,88, 729-745;Spears,K.J.,Roe,A.J.和Gally,D.L.(2006)FEMS Microbiol.Lett.,255,187-202;Liu,B.,Knirel,Y.A.,Feng,L.,Perepelov,A.V.,Senchenkova,S.N.,Wang,Q.,Reeves,P.R.和Wang,L(2008)FEMS Microbiol.Rev.32,627-653;Stenutz,R.,Weintraub,A.和Widmalm,G.(2006)FEMS Microbiol.Rev.30,382-403)。已经鉴定出具有超过180个独立O-血清型(属于独特的O-抗原结构)的大肠杆菌菌株(Stenutz,R.,Weintraub,A.和Widmalm,G.(2006)FEMS Microbiol.Rev.30,382-403)。

[0020] O-抗原重复单位在与十一异戊二烯基焦磷酸连接的内膜细胞溶质面上进行预装配。脂质-连接的重复单位横向地扩散(翻转(flip-flop))到内膜的周质表面并在转运到外膜和连接到LPS之前聚合。大多数杂聚合O-抗原重复单位在还原末端或者具有N-乙酰基葡萄糖胺(“GlcNAc”)或者具有N-乙酰基半乳糖胺(“GalNAc”)。

[0021] 已假定通过被WecA催化的GlcNAc-P或GalNAc-P从其相应的糖核苷酸衍生物转移到十一异戊二烯基单磷酸(“Und-P”)上,启动脂质中间体的生物合成(Samuel,G.和Reeves,P.R.(2003)Carbohydrate Research,338,2503-2519;Alexander,D.C.和Valvano,M.A.(1994)J.Bacteriol.,176,7079-7084;Zhang,L.,Radziejewska-Lebrecht,J.,Krajewska-Pietrasik,D.,Tolvanen,P.和Skurkik,M.(1997)Mol.Microbiol.23,63-76;Amor,P.A.和Whitfield,C.(1997)Mol.Microbiol.26(145-161);Wang,L.和Reeves,P.R.(1998)Infect.Immun.66,3545-3551)。虽然已经表征了WecA的GlcNAc-磷酸转移酶活性的性质和特异性(Rush,J.S.,Rick,P.D.和Waechter,C.J.(1997)Glycobiology,7,315-322),但是WecA催化GalNAc-P-P-Und合成的结论是基于遗传学研究(Wang,L.和Reeves,P.R.(1998)Infect.Immun.66,3545-3551)。这样的较早遗传学研究表明通过被WecA催化的GalNAc-P从UDP-GalNAc酶促转移到Und-P上启动脂质-连接的四糖中间体的生物合成(Wang,L.和Reeves,P.R.(1998)Infect.Immun.66,3545-3551)。然而,没有直接的酶学证据证明WecA利

用UDP-GalNAc作为GalNAc-P供体。

[0022] 此外,先前已提出大肠杆菌055 gne和gne1基因编码UDP-GlcNAc 4-差向异构酶(Wang, L., Huskic, S., Cisterne, A., Rothmund, D. 和Reeves, P. R. (2002) J. Bacteriol. 184, 2620-2625; Guo, H., Yi, W., Li, L. 和Wang, P. G. (2007) Biochem. Biophys. Res. Commun., 356, 604-609)。先前的报导鉴定了分别来自大肠杆菌055 (Wang, L., Huskic, S., Cisterne, A., Rothmund, D. 和Reeves, P. R. (2002) J. Bacteriol. 184, 2620-2625) 和大肠杆菌086 (Guo, H., Yi, W., Li, L. 和Wang, P. G. (2007) Biochem. Biophys. Res. Commun., 356, 604-609) 的两个基因, 分别为大肠杆菌055 gne和大肠杆菌086 gne1, 它们与同一基因家族内的Z3206基因有100%同一性。

[0023] 因此, 将会一直导致技术人员相信Z3206基因也编码UDP-GlcNAc/UDP-GalNAc差向异构酶。

[0024] 发明简述

[0025] 现在令人惊奇地发现, 由大肠杆菌0157中的Z3206基因编码的差向异构酶能催化在十一异戊二烯基焦磷酸上合成N-乙酰基半乳糖胺(“GalNAc”)的反应, 从而启动寡糖或多糖的形成。

[0026] 在一个方面, 本发明涉及一种重组原核生物合成系统, 其产生多糖的全部或一部分, 包含在十一异戊二烯基焦磷酸上合成GalNAc的差向异构酶。本发明进一步包括合成在还原末端具有GalNAc的多糖的全部或一部分的糖基转移酶, 并且再进一步包括合成在还原末端具有GalNAc的抗原性多糖的全部或一部分的糖基转移酶。

[0027] 在另一个方面, 本发明涉及在十一异戊二烯基焦磷酸上产生GalNAc的差向异构酶, 和在一个进一步方面, 该差向异构酶由Z3206基因编码。

[0028] 在一个另外的方面, 本发明涉及用于产生N-糖基化蛋白的表达系统, 包含: 编码寡糖基转移酶的核苷酸序列; 编码蛋白载体的核苷酸序列; 来自至少一种细菌的至少一种寡糖或多糖基因簇; 其中该多糖在还原末端含有GalNAc; 和编码差向异构酶的核酸序列。

[0029] 在一个再进一步方面, 本发明涉及一种包含Z3206基因的重组原核生物合成系统, 该Z3206基因编码将GlcNAc-P-P-Und转变成GalNAc-P-P-Und的差向异构酶。

[0030] 在又一个另外的方面, 本发明涉及一种包含大肠杆菌055 gne基因或大肠杆菌086 gne1基因的重组原核生物合成系统, 该基因编码将GlcNAc-P-P-Und转变成GalNAc-P-P-Und的差向异构酶。

[0031] 在再一个方面, 本发明涉及一种N-糖基化蛋白, 其包含至少一种引入的共有序列D/E-X-N-Z-S/T, 其中X和Z可以为除脯氨酸以外的任何天然氨基酸, 和在还原末端具有N-乙酰基半乳糖胺的聚糖。

[0032] 在还另一个方面, 本发明涉及一种生物缀合物疫苗, 其包含具有至少一种引入的共有序列D/E-X-N-Z-S/T的N-糖基化蛋白, 其中X和Z可以为除脯氨酸以外的任何天然氨基酸; 在还原末端具有N-乙酰基半乳糖胺的免疫原性聚糖; 和佐剂。

[0033] 在一个另外的方面, 本发明涉及在宿主细胞中产生N-联糖基化蛋白的方法, 该宿主细胞包含: 编码由在还原末端含有GalNAc的至少一种细菌装配至少一种寡糖或多糖的糖基转移酶的核酸; 编码蛋白载体的核酸; 编码寡糖基转移酶的核酸; 和编码差向异构酶的核酸。

[0034] 在一个进一步方面,本发明涉及生物合成系统和蛋白质在制备生物缀合物疫苗中的应用。

[0035] 在一个另外的方面,本发明涉及用于产生单糖、寡糖和多糖的方法,和在一个再进一步方面,本发明涉及用于产生抗原性聚糖和N-糖基化蛋白的方法。

[0036] 附图简述

[0037] 图1显示由来自大肠杆菌0157的膜部分合成 $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$ 的时程。将来自大肠杆菌菌株0157的膜部分与 $\text{UDP}-[^3\text{H}]\text{GlcNAc}$ 一起于 37°C 孵育指定时间。按实施例2中描述的方法提取 $[^3\text{H}]$ 脂质产物并测定 $[^3\text{H}]\text{GlcNAc}$ 掺入到 $[^3\text{H}]\text{GlcNAc-P-P-Und}(\text{O})$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}(\bullet)$ 中。

[0038] 图2显示由 GlcNAc-P-P-Und 形成 GalNAc-P-P-Und 的提出的生物合成途径。

[0039] 图3显示由来自大肠杆菌菌株0157的膜部分合成 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的纯化和表征。将来自大肠杆菌0157的膜部分与 $\text{UDP}-[^3\text{H}]\text{GlcNAc}$ 一起孵育,并按实施例3中描述的方法纯化 $[^3\text{H}]\text{GalNAc}$ 脂质。图3A显示在硼酸盐浸渍的硅胶G(Quantum 1)上经DEAE-纤维素纯化后 $[^3\text{H}]\text{HexNAc}$ 脂质的制备型薄层层析图。图3B显示从小图A中制备板中回收后在硼酸盐浸渍的硅胶G(Baker, Si250)上经纯化的 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的薄层层析图。图3C显示在图3B中纯化的 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 经温和的酸水解后回收的 $[^3\text{H}]$ -氨基糖的下行纸层析图(硼酸盐浸渍的瓦特曼1号纸(Whatman No.1 paper))。图3D显示用 NaBH_4 还原来自图3C的 $[^3\text{H}]$ 氨基糖产生的 $[^3\text{H}]\text{HexNAc-糖醇}$ 的下行纸层析图(Whatman No.3MM)。

[0040] 图4显示大肠杆菌21546细胞和大肠杆菌21546细胞在用pMLBAD:Z3206转化后的代谢性标记。大肠杆菌21546(图4A)和大肠杆菌21546:pMLBAD/Z3206(图4B)用 $[^3\text{H}]\text{GlcNAc}$ 于 37°C 进行代谢性标记5分钟。按实施例3中描述的方法提取 $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$,使其不含水溶性污染物后在硼酸盐浸渍的硅胶板(Baker Si250)上通过薄层层析进行分离。使用Bioscan层析扫描仪检测放射性脂质。 GalNAc-P-P-Und 和 GlcNAc-P-P-Und 的层析位置用箭头表示。

[0041] 图5显示由来自大肠杆菌菌株的膜部分与 $\text{UDP}-[^3\text{H}]\text{GlcNAc}$ 一起孵育形成的 $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$ 的薄层层析图。将来自大肠杆菌菌株K12(图5A)、0157(图5B)、21546(图5C)和21546:pMLBAD/Z3206(图5D)的膜部分与 $\text{UDP}-[^3\text{H}]\text{GlcNAc}$ 一起于 37°C 孵育10分钟,按实施例3中描述的方法提取 $[^3\text{H}]$ 脂质产物,通过分配使其不含水溶性污染物,然后在硼酸盐浸渍的硅胶板(Baker Si250)上通过薄层层析进行分离。 GalNAc-P-P-Und 和 GlcNAc-P-P-Und 的层析位置用箭头表示。

[0042] 图6显示 GlcNAc-P 与UMP孵育后的卸载(discharge)。将来自大肠杆菌21546:Z3206的膜部分与 $\text{UDP}-[^3\text{H}]\text{GlcNAc}$ 一起于 37°C 预孵育至将 GlcNAc-P-P-Und 酶标记10分钟(图6A),然后与 1mM UMP一起进行第二个孵育期,包括或 1min (图6B)或 2min (图6C)。在指定的孵育期后,按实施例3中描述的方法提取 $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$,并在硼酸盐浸渍的硅胶板(Baker Si250)上通过薄层层析进行解析。 GalNAc-P-P-Und 和 GlcNAc-P-P-Und 的层析位置用箭头表示。

[0043] 图7显示由来自表达Z3206的菌株21546的膜催化的外源 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 转变成直接有关的 $[^3\text{H}]\text{HexNAc-P-P-Und}$ 产物。将来自大肠杆菌菌株21546(图7B和图7E)和21546:pMLBAD/Z3206(图7C和图7F)的膜部分与纯化的 $[^3\text{H}]\text{GlcNAc-}$

P-P-Und(图7A,图7B和图7C)或 $[^3\text{H}]\text{GalNAc-P-P-Und}$ (图7D,图7E和图7F的分小图)(在1%曲通X-100(Triton X-100)中超声处理进行分散)一起于37℃孵育1分钟。按实施例3中描述的方法提取 $[^3\text{H}]\text{GlcNAc/GalNAc-P-P-Und}$,在硼酸盐浸渍的硅胶板(Baker Si250)上通过薄层层析进行解析并用Bioscan AR2000放射性层析扫描仪进行检测。

[0044] 图8显示未糖基化和糖基化AcrA蛋白的SDS-PAGE分析。由携带AcrA表达质粒和与pMLBAD:Z3206(第1泳道)、pMLBAD:gne(第2泳道)或载体对照pMLBAD(第3泳道)互补的pgl操纵子 Δ gne的大肠杆菌DH5 α 细胞制备的周质提取物用10%SDS-PAGE分离后转移到硝化纤维素膜上。用抗AcrA抗血清检测AcrA及其糖基化形式。显示了对应于未糖基化(AcrA)和糖基化AcrA(gAcrA)的条带位置。

[0045] 图9显示已经由Liu B等人鉴定的基因(Liu B等人,Structure and genetics of Shigella O antigens(志贺氏菌O抗原的结构和遗传学)FEMS Microbiology Review, 2008.32:p.27)。

[0046] 图10是显示含有合成弗氏志贺氏菌6(*S.flexneri* 6)O抗原所需基因的DNA区的方案。

[0047] 图11显示弗氏志贺氏菌6 O抗原在大肠杆菌中的表达。LPS通过银染色或者通过转移到硝化纤维素膜和通过抗弗氏志贺氏菌6的抗体检测予以显现。

[0048] 图12显示O抗原的HPLC。含有弗氏志贺氏菌-Z3206的大肠杆菌细胞(SCM3)、含有弗氏志贺氏菌+Z3206的大肠杆菌细胞(SCM3)或空大肠杆菌(SCM3)细胞的LL0分析。

[0049] 图13显示来自表达EPA、pglB和弗氏志贺氏菌6O-抗原+/-Z3206的大肠杆菌细胞的经镍纯化的蛋白的蛋白质印迹图。

[0050] 发明详述

[0051] 本发明包括一种重组原核生物合成系统,其包含编码差向异构酶的核酸,该差向异构酶合成在还原末端具有N-乙酰基半乳糖胺的寡糖或多糖和在聚糖的还原末端具有N-乙酰基半乳糖胺的N-糖基化蛋白。

[0052] 术语“部分氨基酸序列”也称为“最优化共有序列”或“共有序列”。最优化共有序列被寡糖基转移酶(“OST,”“OTase”)进行N-糖基化,其效率比常规共有序列“N-X-S/T”高得多。

[0053] 一般而言,术语“重组N-糖基化蛋白”是指在天然不包含所述蛋白编码核酸的宿主细胞中产生的任何多肽或寡肽。在本发明的正文中,该术语是指在原核宿主细胞例如埃希氏菌(*Escherichia* spp.)、弯曲杆菌(*Campylobacter* spp.)、沙门氏菌(*Salmonella* spp.)、志贺氏菌(*Shigella* spp.)、螺杆菌(*Helicobacter* spp.)、假单胞菌(*Pseudomonas* spp.)、芽孢杆菌(*Bacillus* spp.)中,和在进一步实施方案大肠杆菌(*Escherichia coli*)、空肠弯曲杆菌(*Campylobacter jejuni*)、鼠伤寒沙门氏菌(*Salmonella typhimurium*)等中重组产生的蛋白质,其中将所述蛋白编码核酸导入到所述宿主细胞中并且其中所编码的蛋白被OTase进行N-糖基化,所述转移酶天然存在于或者被重组地导入到所述宿主细胞中。

[0054] 根据氨基酸的国际上接受的单字母代码,缩写词D、E、N、S和T分别表示天冬氨酸、谷氨酸、天冬酰胺、丝氨酸和苏氨酸。

[0055] 根据本发明所述的蛋白质包含被引入到蛋白质并被N-糖基化的最优化共有序列D/E-X-N-Z-S/T中的一个或多个。因此,本发明的蛋白质不同于也含有该最优化共有序列但

不包含任何额外的(引入的)最优化共有序列的天然存在的空肠弯曲杆菌N-糖蛋白。

[0056] 最优化共有序列的引入可以通过一个或多个氨基酸的添加、缺失和/或取代来完成。可通过化学合成策略(例如固相辅助的化学肽合成)完成一个或多个氨基酸的添加、缺失和/或取代以引入最优化共有序列,根据本发明,这项技术对于本领域技术人员来说是众所周知的。对于较大的多肽来说可选和优选的是,根据本发明,可通过作为本领域标准技术的重组技术来制备本发明的蛋白质。

[0057] 本发明的蛋白质具有它们可以以高效和在任何宿主中产生的优点。在本发明的一个实施方案中,宿主包含来自弯曲杆菌、例如来自空肠弯曲杆菌的功能性pgl操纵子。在进一步实施方案中,来自弯曲杆菌并用于实施本发明的寡糖基转移酶是来自大肠弯曲杆菌(*Campylobacter coli*)或海鸥弯曲杆菌(*Campylobacter lari*)。根据本发明,寡糖基转移酶对于本领域技术人员来说是显而易见的。例如,寡糖基转移酶公开于诸如以下的参考文献中:Szymanski, C.M. 和 Wren, B.W. (2005) Protein glycosylation in bacterial mucosal pathogens (细菌粘膜病原体中的蛋白质糖基化), *Nat. Rev. Microbiol.* 3:225-237。当所述原核宿主是弯曲杆菌,或者例如空肠弯曲杆菌时,功能性pgl操纵子可天然存在。然而,根据本领域之前所证明的及以上提及的,pgl操纵子可以被转移到细胞中并在所述新的细胞环境中仍保留有功能。

[0058] 术语“来自弯曲杆菌、优选空肠弯曲杆菌的功能性pgl操纵子”意指编码弯曲杆菌、例如空肠弯曲杆菌的功能性寡糖基转移酶(OTase)和能够在脂质载体上装配寡糖的一个或多个特异性糖基转移酶的核酸簇,并且其中所述寡糖可以通过OTase从脂质载体转移到具有一个或多个最优化氨基酸序列D/E-X N-Z-S/T的靶蛋白上。应当理解,术语“来自弯曲杆菌、优选空肠弯曲杆菌的功能性pgl操纵子”在本发明的正文中不一定就是指作为一个转录单位的操纵子。该术语只是要求在一个宿主细胞中重组蛋白N-糖基化的功能性组分存在。这些组分可以转录成为一个或多个独立的mRNA并且可以被一起或单独地调节。例如,该术语也包括位于一个宿主细胞中基因组DNA和质粒上的功能性组分。为了效率,在一个实施方案中,功能性pgl操纵子的所有组分都可以被同时调节和表达。

[0059] 寡糖基转移酶,在有些实施方案中可来源于弯曲杆菌,而在其它实施方案中可来源于空肠弯曲杆菌。在另外的实施方案中,寡糖基转移酶可来源于本领域技术人员已知的具有寡糖基转移酶的生物体,例如沃林氏菌(*Wolinella* spp.)和真核生物。

[0060] 能够在脂质载体上装配寡糖的一个或多个特异性糖基转移酶可来源于宿主细胞或被重组地引入到所述宿主细胞中,唯一的限制性是由所述糖基转移酶装配的寡糖可以通过OTase从脂质载体转移到具有一个或多个最优化共有序列的靶蛋白上。因此,天然包含特异性糖基转移酶的宿主细胞的选择和/或替代所述宿主中天然存在的特异性糖基转移酶以及异源特异性糖基转移酶的引入都将能够使本领域技术人员去改变与本发明蛋白质中的最优化N-糖基化共有位点结合的N-聚糖。

[0061] 根据以上结果,本发明提供本发明蛋白质上N-聚糖模式的个体设计。因此,所述蛋白在它们的N-聚糖模式上可以是个性化的以满足生物学、药学和纯化的要求。

[0062] 在本发明的实施方案中,所述蛋白可包含所述N-糖基化最优化氨基酸序列中的一个但也可不止一个,例如至少两个、至少3个或至少5个。

[0063] 本发明蛋白质中一个或多个N-糖基化最优化氨基酸序列的存在可具有增加其免

疫原性、增加其稳定性、影响其生物活性、延长其生物半寿期和/或使其纯化简化的优点。

[0064] 最优化共有序列在位置X和Z上可包括除脯氨酸以外的任何氨基酸。术语“任何氨基酸”意指包括常用和稀有天然氨基酸以及合成氨基酸衍生物和类似物，它们都将仍然允许最优化共有序列被OTase进行N-糖基化。天然存在的常用和稀有氨基酸对于X和Z是优选的。X和Z可以相同或不同。

[0065] 需要注意的是，对于根据本发明所述的蛋白质中的每个最优化共有序列，X和Z可以是不同的。

[0066] 可以通过特异性糖基转移酶及其当在脂质载体上通过OTase进行转移来装配寡糖时相互作用来决定与最优化共有序列结合的N-聚糖。根据本发明，本领域技术人员通过改变所需宿主细胞中存在的特异性糖基转移酶的类型和数量将能够设计出N-聚糖。

[0067] 本文所用的“单糖”是指一个糖残基。“寡糖和多糖”是指两个或更多个糖残基。本文所用的术语“聚糖”是指单糖、寡糖或多糖。“N-聚糖”在本文中定义为蛋白质中通过N-糖苷键与天冬酰胺残基的ε-酰胺氮连接的可变组成的单糖、寡糖或多糖。在一个实施方案中，通过OTase转移的N-聚糖在革兰氏阴性或阳性细菌的细胞质膜中存在的十一异戊烯醇焦磷酸(“Und-P-P”)脂质-锚形体上进行装配。它们参与O抗原、O多糖和肽聚糖的合成(Bugg, T.D.和Brandish, P.E.(1994). From peptidoglycan to glycoproteins: common features of lipid-linked oligosaccharide biosynthesis(从肽聚糖到糖蛋白: 脂质连接的寡糖生物合成的共同特征). FEMS Microbiol Lett 119, 255-262; Valvano, M.A. (2003). Export of O-specific lipopolysaccharide(O-特异性脂多糖的输出). Front Biosci 8, s452-471)。

[0068] 进行了多项研究以决定脂质连接的重复四糖(4-N-乙酰基perosamine→岩藻糖→葡萄糖→GalNAc)的生物合成是否是通过WecA形成GalNAc-P-P-Und启动的。当来自大肠杆菌菌株K12、0157和PR4019(一种WecA-过量表达菌株)的膜部分与UDP-[³H]GalNAc一起孵育时，既未检测到[³H]GlcNAc-P-P-Und也未检测到[³H]GalNAc-P-P-Und的酶促合成。然而，当来自菌株0157的膜部分与UDP-[³H]GlcNAc一起孵育时，观察到两种酶标记的产物具有[³H]GlcNAc-P-P-Und和[³H]GalNAc-P-P-Und的化学和层析特性，证实菌株0157含有能够使GlcNAc-P-P-Und和GalNAc-P-P-Und相互转变的差向异构酶。当外源[³H]GlcNAc-P-P-Und与来自菌株0157的膜一起孵育时，外源[³H]GlcNAc-P-P-Und被转变成[³H]GalNAc-P-P-Und也证实了差向异构酶的存在。当菌株0157用[³H]GlcNAc进行代谢性标记时，[³H]GlcNAc-P-P-Und和[³H]GalNAc-P-P-Und两者都被检出。用Z3206基因转化大肠杆菌菌株21546能够使这些细胞在体内(in vivo)和体外(in vitro)合成GalNAc-P-P-Und。当来自菌株0157的膜与外源[³H]GalNAc-P-P-Und一起孵育时重新形成[³H]GlcNAc-P-P-Und，证明了差向异构酶反应的可逆性。在空肠弯曲杆菌N-糖基化系统在大肠杆菌中的表达中，Z3206不能代偿gne基因的丢失，表明它没有作为UDP-GlcNAc/UDP-GalNAc差向异构酶起作用。基于这些结果，证实了在大肠杆菌0157中，GalNAc-P-P-Und通过GlcNAc-P-P-Und差向异构酶可逆地合成后通过WecA形成GlcNAc-P-P-Und。

[0069] 对大肠杆菌0157 O-抗原亚基装配的起始反应进行研究证实GalNAc-P-P-Und合成是被有些先前未知的机制而不是被WecA催化的。本文给出的证据说明在大肠杆菌0157中，GalNAc-P-P-Und不是由WecA催化的GalNAc-P从UDP-GalNAc转移合成的，而是由Z3206基因

编码的差向异构酶催化的GlcNAc-P-P-Und的4-OH可逆差向异构化合成的。

[0070] 因此,本发明包括用于装配重要的细菌细胞表面组分的一种新的生物合成途径以及用于合成GalNAc-P-P-Und的一种新的生物合成途径。本发明的一个进一步实施方案包括作为抗微生物剂新靶标的细菌差向异构酶。

[0071] 大肠杆菌0157合成具有重复四糖结构(4-N-乙酰基perosamine→岩藻糖→葡萄糖→GalNAc)的O-抗原。本文说明了脂质连接的四糖中间体的生物合成不是通过由WecA催化的GalNAc-P从UDP-GalNAc酶促转移到Und-P而启动的,这与较早的遗传学研究正好相反(Wang, L. 和Reeves, P. R. (1998) *Infect. Immun.* 66, 3545-3551)。本文描述的发明是通过同源性检索获得后再通过遗传学、酶学和代谢性标记实验的结果来验证的,证明WecA没有利用UDP-GalNAc作为底物,而是需要WecA来合成GlcNAc-P-P-Und,然后通过菌株0157中Z3206基因编码的差向异构酶将其可逆地转变成GalNAc-P-P-Und。

[0072] 本发明的Z3206基因属于产生在其还原末端含有GalNAc残基的表面O-抗原重复单位的几个菌株中存在的基因家族(表1)。Z3206基因序列示于SEQ ID NO:1。先前的报导鉴定了分别来自大肠杆菌055(Wang, L., Huskic, S., Cisterne, A., Rothmund, D. 和Reeves, P. R. (2002) *J. Bacteriol.* 184, 2620-2625)和大肠杆菌086(Guo, H., Yi, W., Li, L. 和Wang, P. G. (2007) *Biochem. Biophys. Res. Comm.*, 356, 604-609)的两个基因即大肠杆菌055gne和大肠杆菌086 gne1,这两个基因与Z3206基因有100%同一性(表1)。大肠杆菌055gne基因序列显示为SEQ ID NO:3,而大肠杆菌086 gne1基因序列显示为SEQ ID NO:5。

[0073] 表1

[0074] Z3206基因在表达在还原末端具有GalNAc的O-抗原链的细菌菌株中的相关性

[0075]

	与 Z3206 的%同一性	在 O-抗原重复单位的还原末端的 GalNAc
大肠杆菌 O55 <i>gne</i> (SEQ ID NO: 3)	100	有
大肠杆菌 O86 <i>gne1</i> (SEQ ID NO: 5)	100	有
鲍氏志贺氏菌(<i>Shigella boydii</i>) O18 <i>gne</i> (SEQ ID NO: 7)	88	有
肠沙门氏菌(<i>Salmonella enteric</i>) O30 <i>gne</i> (SEQ ID NO: 9)	94	有
空肠弯曲杆菌 <i>gne</i> (SEQ ID NO: 11)	21	无
大肠杆菌 K12 <i>galE</i> (SEQ ID NO: 13)	27	无
大肠杆菌 O86 <i>gne2</i> (SEQ ID NO: 15)	18	有

[0076] 因此,我们得出的结论是大肠杆菌055gne和大肠杆菌086gne1分别在菌株055和086中也编码能够将GlcNAc-P-P-Und转变成GalNAc-P-P-Und的差向异构酶,这也产生了在还原末端具有GalNAc的O-抗原重复单位(表1)。

[0077] 在本研究中的两个实验方法表明Z3206蛋白不催化菌株0157中UDP-GlcNAc向UDP-GalNAc的差向异构化。首先,当将来自菌株0157的膜与³H]UDP-GalNAc一起孵育时,既未检测到³H]GlcNAc-P-P-Und也未检测到³H]GalNAc-P-P-Und(表3)。如果Z3206催化³H]UDP-GalNAc向³H]UDP-GlcNAc的转变,将预期观察到³H]GlcNAc-P-P-Und。其次,我们已经表明

血凝素标记的Z3206不能代偿依赖UDP-GalNAc的空肠弯曲杆菌N-糖基化报道系统(图8)。

[0078] 来自菌株055的大肠杆菌055gne基因(Wang, L., Huskic, S., Cisterne, A., Rothmund, D.和Reeves, P.R. (2002) J. Bacteriol. 184, 2620-2625)也测定了差向异构酶活性,即将粗提取物与UDP-GalNAc一起孵育并通过测定酸水解后与对二甲氨基苯甲醛的反应性增加来间接测定向UDP-GlcNAc的转变。在两个研究中,产物的形成是基于与对二甲氨基苯甲醛的反应性变化并且不是糖核苷酸终产物的决定性特征。在偶联测定中,相对于Gne2, 90%纯的多聚组氨酸标记的大肠杆菌086 gne1也显示具有低水平的UDP-葡萄糖差向异构酶活性。

[0079] 因此,本发明的一个实施方案涉及一种重组原核生物合成系统,其含有将GlcNAc-P-P-Und转变成GalNAc-P-P-Und的Z3206基因、大肠杆菌055 gne基因或大肠杆菌086 gne1基因。

[0080] 显然,大肠杆菌086(其合成含有两个GalNAc残基的O-抗原,其假定需要UDP-GalNAc作为额外非还原末端GalNAc的糖基供体),也具有额外的GlcNAc 4-差向异构酶基因,称为gne2,在O-抗原基因簇内(Guo, H., Yi, W., Li, L.和Wang, P.G. (2007) Biochem. Biophys. Res. Commun., 356, 604-609)。这个额外的差向异构酶基因与荚膜异多糖酸(colanic acid)基因簇的galE基因有高同源性并且似乎是能够合成UDP-GalNAc的UDP-GlcNAc 4-差向异构酶。

[0081] Z3206基因似乎在用GalNAc起始的大肠杆菌O-血清型中是高度保守的。在最近的一项研究中,通过基于聚合酶链式反应的方法,利用针对特异性检测大肠杆菌0157 Z3206基因设计的核苷酸引物,根据Z3206的表达筛选出62个具有建立的O-抗原重复单位结构的大肠杆菌菌株(Wang, L., Huskic, S., Cisterne, A., Rothmund, D.和Reeves, P.R. (2002) J. Bacteriol. 184, 2620-2625)。在该项研究中,在已知含有GalNAc的22个大肠杆菌菌株中有16个检测到Z3206,而在缺乏GalNAc的40个菌株中只有4个检测到Z3206。而且,用设计检测具有UDP-GlcNAc 4-差向异构酶活性的替代差向异构酶的引物类似筛选22个含GalNAc的菌株(大肠杆菌0113的galE基因)没有检测到携带该基因的菌株,表明Z3206是大肠杆菌的O-抗原重复单位中与还原末端GalNAc的存在最常见相关的GlcNAc 4-差向异构酶基因。

[0082] 通过多种基于web的拓扑预测算法分析Z3206蛋白质序列表明Z3206蛋白不是高度疏水性的。绝大多数拓扑预测算法表明Z3206是一种可溶性37kDa蛋白,尽管TMPred (Hofmann, K.和Stoffel, W. (1993) Biol. Chem. Hoppe-Seyler 374, 166(摘要))预测有一个弱的N-端跨膜螺旋。然而,来自表达血凝素标记的Z3206的大肠杆菌细胞的细胞馏份的SDS-PAGE后的蛋白质转印清楚地显示标记的蛋白与细胞低渗裂解后的颗粒部分缔合。初步实验表明该蛋白质在膜部分与1MKCl孵育后,仍保留与颗粒部分缔合,但是与0.1%Triton X-100孵育呈有活性的形式增溶。

[0083] 大肠杆菌0157Z3206与氧化还原酶的短链脱氢酶/还原酶家族具有显著序列同源性,包括GXXGXXG基序(罗斯曼折叠(Rossman fold)),与NAD(P)结合口袋(Allard, S.T.M., Giraud, M.F.和Naismith, J.H. (2001) Cell. Mol. Life Sci. 58, 1650-1655)和保守的S X₂4YX₃K序列相一致,参与质子夺取和供给(Field, R.A.和Naismith, J.H. (2003) Biochemistry 42, 7637-7647)。基于UDP-Glc 4-差向异构酶即短链脱氢酶/还原酶家族的另一个成员的晶体结构的分子建模,提示在氢化物夺取之后,4-酮基中间体围绕UDP的β磷

酸旋转以呈现该酮基中间体的相反一面并允许从该相反一侧重插入氢化物,因此在碳4上羟基的构型发生颠倒。这些保守序列的存在提示Z3206很可能通过类似的机制起作用。虽然图7中见到的差向异构酶产物的平衡分布好像有利于GlcNAc-P-P-Und的形成,但是利用GalNAc-P-P-Und进行O-抗原重复单位装配将按GalNAc-P-P-Und的方向通过质量作用驱动差向异构化反应。

[0084] 自然界中聚类异戊二烯脂质中间体的糖基部分的差向异构化一直未广泛报道。在一个先前的研究中,报道了核糖基-P-十异戊烯醇的2-差向异构化形成阿拉伯糖基-P-十异戊烯醇,分枝杆菌中阿拉伯半乳聚糖生物合成的一种阿拉伯糖基供体(Mikusová, K., Huang, H., Yagi, T., Holsters, M., Vereecke, D., D'Haese, W., Scherman, M. S., Brennan, P. J., McNeil, M. R. 和 Crick, D. C. (2005) *J. Bacteriol.* 187, 8020-8025)。阿拉伯糖基-P-十异戊烯醇通过两步氧化/还原反应来形成,需要Rv3790和Rv3791两个分枝杆菌蛋白。虽然加入NAD和NADP能适度地刺激差向异构化,但是Rv3790和Rv3791两者都不含有罗斯曼折叠或SX₂₄YXXXX基序,即短链脱氢酶/还原酶家族的特征(Allard, S. T. M., Giraud, M.-F. 和 Naismith, J. H. (2001) *Cell. Mol. Life Sci.* 58, 1650-1655; Field, R. A. 和 Naismith, J. H. (2003) *Biochemistry* 42, 7637-7647)。

[0085] 总之,描述了通过GlcNAc-P-P-Und差向异构化形成GalNAc-P-P-Und的一种新的生物合成途径。

[0086] 已经表明有若干抗生素能抑制GlcNAc-P-P-Und的合成,但是在其应用中有限制,因为它们也阻断GlcNAc-P-P-多萜醇(蛋白质N-糖基化途径的起始多萜醇连接的中间体)的合成。虽然GlcNAc-P-P-多萜醇是细菌糖脂中间体GlcNAc-P-P-Und的一种结构上相关的哺乳动物对应物(counterpart),但是没有关于在真核细胞中将GlcNAc-P-P-多萜醇转变成GalNAc-P-P-多萜醇的类似差向异构化反应的证据。因此,这提高了在其中在还原末端含有GalNAc的表面O-抗原参与病理过程的菌株中,O-抗原合成可以潜在地通过抑制细菌差向异构酶来阻断的可能性。

[0087] 本发明的一个实施方案涉及在大肠杆菌0157中将GlcNAc-P-P-Und(N-乙酰基葡萄糖胺基焦磷酸基十一异戊烯醇)转变成GalNAc-P-P-Und(N-乙酰基半乳糖胺基焦磷酸基十一异戊烯醇)的差向异构酶。本发明的一个再进一步的示例性方面涉及在还原末端具有GalNAc的脂质结合的重复四糖合成的起始。

[0088] 本发明另一个方面的基础包括发现了空肠弯曲杆菌含有通用的N-联蛋白质糖基化系统。已经表明空肠弯曲杆菌的各种蛋白都被七糖进行了修饰。这种七糖在十一异戊二烯基焦磷酸即载体脂质上在内膜的细胞质一侧进行装配,即通过逐步加入特异性糖基转移酶催化的核苷酸活化单糖。然后,该脂质连接的寡糖通过翻转酶(例如Pg1K)翻转(横向地扩散)到周质间隙。在N-联蛋白质糖基化的最后一步,寡糖基转移酶(例如Pg1B)催化寡糖从载体脂质转移到具有共有序列D/E-X-N-Z-S/T的天冬酰胺(Asn)残基上,其中X和Z可以为除Pro以外的任何氨基酸。七糖的糖基化簇已成功地转移到大肠杆菌中并产生了弯曲杆菌(*Campylobacter*)的N-联糖蛋白。

[0089] 已经证明Pg1B对于脂质连接的糖底物没有严格的特异性。在十一异戊二烯基焦磷酸上装配的抗原性多糖通过周质中的Pg1B俘获并转移到蛋白载体上(Feldman, 2005; Wacker, M. 等人, Substrate specificity of bacterial oligosaccharyltransferase

suggests a common transfer mechanism for the bacterial and eukaryotic systems (细菌寡糖基转移酶的底物特异性提示细菌和真核系统的一种共同转移机制). Proc Natl Acad Sci U S A, 2006. 103(18): p. 7088-93。该酶也将转移各式各样的十一异戊二烯基焦磷酸(UPP)连接的寡糖, 如果它们在还原末端含有N-乙酰基化己糖胺的话。在W02009/104074中公布了pglB的核苷酸序列和PglB的氨基酸序列。

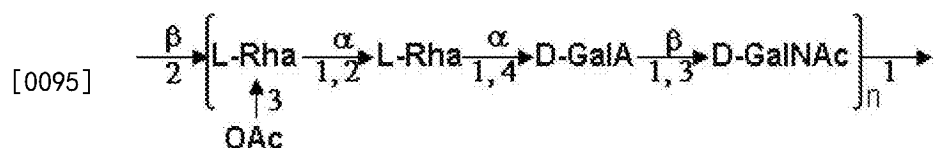
[0090] 因此, 本发明的一个实施方案涉及一种重组N-糖基化蛋白, 其包含: 引入的共有序列D/E-X-N-Z-S/T中的一个或多个, 其中X和Z可以为除脯氨酸以外的任何天然氨基酸; 和在还原末端具有N-乙酰基半乳糖胺和并且与所述一个或多个引入的共有序列的每一个通过N-糖苷键N-连接的寡糖或多糖。

[0091] 在一个进一步实施方案中, 本发明涉及一种用于产生多糖的全部或一部分的重组原核生物合成系统, 其包含在十一异戊二烯基焦磷酸上合成N-乙酰基半乳糖胺(“GalNAc”)的差向异构酶。在一个进一步实施方案中, 多糖的全部或一部分是有抗原性的。

[0092] 在另一个实施方案中, 本发明涉及一种重组原核生物合成系统, 其包含: 在十一异戊二烯基焦磷酸上合成GalNAc的差向异构酶; 和合成在还原末端具有GalNAc的多糖的糖基转移酶。

[0093] 本发明的一个实施方案进一步包含一种重组原核生物合成系统, 其包含在十一异戊二烯基焦磷酸上合成GalNAc的差向异构酶和合成多糖的糖基转移酶, 其中所述多糖具有以下结构: α -D-PerNAc- α -L-Fuc- β -D-Glc- α -D-GalNAc; 并且其中GalNAc是在所述多糖的还原末端。

[0094] 重组原核生物合成系统可以产生各种来源的单糖、寡糖或多糖。本发明的实施方案涉及各种来源的寡糖和多糖。这样的寡糖和多糖可以是原核或真核生物来源的。原核生物来源的寡糖或多糖可以来自革兰氏阴性或革兰氏阳性细菌。在本发明的一个实施方案中, 寡糖或多糖是来自大肠杆菌。在本发明的一个进一步方面中, 所述寡糖或多糖是来自大肠杆菌0157。在另一个实施方案中, 所述寡糖或多糖包含以下结构: α -D-PerNAc- α -L-Fuc- β -D-Glc- α -D-GalNAc。在本发明的一个进一步实施方案中, 寡糖或多糖是来自弗氏志贺氏菌。在一个再进一步实施方案中, 寡糖或多糖是来自弗氏志贺氏菌6。在一个再进一步方面, 所述寡糖或多糖包含以下结构:



[0096] 本发明的实施方案进一步包括各种来源的蛋白质。这样的蛋白质包括对原核和真核生物来说是天然的蛋白质。蛋白载体可以是例如AcrA或者经修饰含有蛋白质糖基化的共有序列即D/E-X-N-Z-S/T的蛋白载体, 其中X和Z可以为除脯氨酸以外的任何氨基酸(例如修饰的外毒素铜绿假单胞菌(*Pseudomonas aeruginosa*)(“EPA”))。在本发明的一个实施方案中, 蛋白质是铜绿假单胞菌EPA。

[0097] 本发明的一个进一步方面涉及在N-聚糖的还原末端具有GalNAc的新的生物缀合物疫苗。本发明的一个另外的实施方案涉及用于产生这类生物缀合物疫苗的一种新方法, 该方法使用含有在十一异戊二烯基焦磷酸上产生GalNAc的差向异构酶的重组细菌细胞。在一个实施方案中, 生物缀合物疫苗可以用来治疗或预防细菌疾病。在进一步实施方案中, 生物

缀合物疫苗可具有治疗和/或预防癌症或其它疾病的潜力。

[0098] 用于人体的典型接种剂量为约1-25 μ g,优选约1 μ g至约10 μ g,最优选约10 μ g。任选地,疫苗,例如本发明的生物缀合物疫苗包括佐剂。

[0099] 在一个另外的实施方案中,本发明涉及用于产生抵抗至少一种细菌的生物缀合物疫苗的表达式,其包含:编码寡糖基转移酶的核苷酸序列;编码蛋白载体的核苷酸序列;来自至少一种细菌的至少一种多糖基因簇,其中该多糖在还原末端含有GalNAc;和编码差向异构酶的核酸序列。在一个进一步实施方案中,多糖基因簇编码抗原性多糖。

[0100] 在又一个进一步实施方案中,本发明涉及用于产生抵抗至少一种细菌的生物缀合物疫苗的表达式,其包含:编码寡糖基转移酶的核苷酸序列;编码包含至少一种插入的共有序列D/E-X-N-Z-S/T的蛋白载体的核苷酸序列,其中X和Z可以为除脯氨酸以外的任何天然氨基酸;来自至少一种细菌的至少一种多糖基因簇,其中该多糖在还原末端含有GalNAc;和Z3206基因。在一个进一步实施方案中,多糖基因簇编码抗原性多糖。

[0101] 在又一个实施方案中,本发明涉及一种生物缀合物疫苗,其包含:蛋白载体;至少一种与蛋白载体连接的免疫原性多糖链,其中所述多糖在还原末端具有GalNAc,和进一步其中所述GalNAc直接与蛋白载体连接;和佐剂。

[0102] 在又一个另外的实施方案中,本发明涉及一种生物缀合物疫苗,其包含:包含至少一种插入的共有序列D/E-X-N-Z-S/T的蛋白载体,其中X和Z可以为除脯氨酸以外的任何天然氨基酸;来自至少一种细菌并与蛋白载体连接的至少一种免疫原性多糖,其中该至少一种免疫原性多糖在还原末端含有直接与蛋白载体连接的GalNAc;和任选地,佐剂。

[0103] 本发明的另一个实施方案涉及产生生物缀合物疫苗的方法,所述方法包括:在重组生物体中通过使用糖基转移酶装配在还原末端具有GalNAc的多糖;在所述重组生物体中将所述GalNAc与一个或多个靶蛋白的天冬酰胺残基相连接,其中所述一个或多个靶蛋白含有一个或多个T-细胞表位。

[0104] 在一个进一步实施方案中,本发明涉及产生生物缀合物疫苗的方法,所述方法包括:将编码执行靶蛋白N-糖基化的代谢装置(apparatus)的遗传信息导入原核生物中,产生修饰的原核生物;其中将一个或多个重组靶蛋白表达所需的遗传信息导入所述原核生物中;其中将大肠杆菌菌株0157差向异构酶表达所需的遗传信息导入所述原核生物中;和其中该代谢装置包含将在还原末端具有GalNAc的多糖装配在脂质载体上的类型的糖基转移酶,和寡糖基转移酶,该寡糖基转移酶将多糖的GalNAc与靶蛋白的天冬酰胺残基共价连接,和含有至少一种T-细胞表位的靶蛋白;产生经修饰的原核生物的培养物;和从培养基中获得糖基化蛋白。

[0105] 本发明的一个进一步方面涉及一种药物组合物。本发明的一个另外的方面涉及一种药物组合物,其包含根据本发明所述的至少一种N-糖基化蛋白。根据本文的公开内容,包含蛋白的药物的制备在本领域是众所周知的。本发明的一个再进一步方面涉及一种药物组合物,其包含抑制将GlcNAc-P-P-Und转变成GalNAc-P-P-Und的差向异构酶的抗生素。在一个优选的实施方案中,本发明的药物组合物包含药学上可接的赋形剂、稀释剂和/或佐剂。

[0106] 合适的赋形剂、稀释剂和/或佐剂在本领域是众所周知的。赋形剂或稀释剂可以是固体、半固体或可用作活性成分的溶媒或介质的液体材料。本领域普通技术人员在制备组合物的领域中根据所选产品的特殊性质、待治疗疾病或病症、疾病或病症的阶段以及其它

相关情况,可以容易地选出适当的给药形式和给药方式(Remington's Pharmaceutical Sciences, Mack Publishing Co.(1990))。药学上可接的稀释剂或赋形剂的比例和性质由所选药用活性化合物的溶解度和化学特性、精选的给药途径及标准药学实践来决定。药物制剂可适合于口服、胃肠外或局部使用并且可以按片剂、胶囊剂、栓剂、溶液剂、混悬剂等形式给予患者。本发明的药用活性化合物当它们本身有效时,为了稳定性、结晶方便、溶解度增加等目的,可以进行配制并按其药学上可接的盐(例如酸加成盐或碱加成盐)的形式给予。

[0107] 在标注特定核苷酸或氨基酸序列的情况下,应当理解,本发明包括同源序列,该同源序列仍然执行与所标注序列相同的功能性。在本发明的一个实施方案中,这样的序列有至少85%同源性。在另一个实施方案中,这样的序列有至少90%同源性。在再进一步实施方案中,这样的序列有至少95%同源性。

[0108] 两个核苷酸或氨基酸序列间百分同一性的测定方法是本领域技术人员已知的。

[0109] 本文描述的核酸序列,例如以下序列表描述的核酸序列都仅仅是实例,对本领域技术人员来说显而易见的是,这些序列可以按不同的方式进行组合。本发明另外的实施方案包括核酸的变异体。核酸的变异体(例如密码子优化核酸)可以是实质上相同的,即与SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:13、SEQ ID NO:15、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20、SEQ ID NO:21、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:27、SEQ ID NO:28或SEQ ID NO:29有至少80%同一性,例如80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.5%同一性。含有SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:13、SEQ ID NO:15、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20、SEQ ID NO:21、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:27、SEQ ID NO:28或SEQ ID NO:29的序列的核酸变异体包括相对含有SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:7、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:13、SEQ ID NO:15、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20、SEQ ID NO:21、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:27、SEQ ID NO:28或SEQ ID NO:29或其部分的序列来说具有一个或多个核苷酸(例如2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、125、150、175或200个核苷酸)的取代、变异、修饰、置换、缺失和/或添加的核酸。

[0110] 例如,在本发明的一个实施方案中,这样的变异体包括编码将GlcNAc-P-P-Und转变成GalNAc-P-P-Und的差向异构酶并且i)在宿主细胞例如大肠杆菌中表达和ii)与SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7或SEQ ID NO:9或其部分实质上相同的核酸。

[0111] 本文描述的核酸包括重组DNA和合成(例如化学合成)DNA。核酸可以是双链或单链的。就单链核酸而论,核酸可以是有义链或反义链。核酸可以利用寡核苷酸类似物或衍生物来合成。

[0112] 包括本文所述核酸的质粒可以转染或转化到宿主细胞中进行表达。用于转染和转化的技术是本领域技术人员已知的。

[0113] 本文提及的所有出版物都通过引用其全部作为参考。应当理解,本文所用的术语

“或”表示可选择,适当的时候可以组合;也就是说,术语“或”包括所列举的独立选择的每一个以及它们的组合。除非文中清楚地描绘,否则本文所用的单数术语包括它们的复数形式,并且提及复数时包括单数。

[0114] 参照以下实施例对本发明作了进一步限定,以下实施例进一步描述了本发明的组合物和方法,以及它的实用性。对本领域技术人员显而易见的是,在本发明的范围内可以实施对组合物和方法两者的修改。

实施例

[0115] 细菌菌株和质粒-大肠杆菌菌株PR4019(Rush, J.S., Rick, P.D. 和Waechter, C.J. (1997) *Glycobiology*, 7, 315-322)和PR21546(Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. 和Rick, P.D. (1990) *J. Biol. Chem.*, 265, 13490-13497)是由Paul Rick博士(Bethesda, MD)慷慨馈赠的礼物,而大肠杆菌0157:H45(Stephan, R., Borel, N., Zweifel, C., Blanco, M. 和Blanco, J.E. (2004) *BMC Microbiol* 4:10)是得自Claudio Zweifel博士(苏黎世大学兽医研究所(Veterinary Institute, University of Zurich))的礼物。使用大肠杆菌DH5 α (Invitrogen)作为克隆实验和蛋白质糖基化分析的宿主。使用的质粒列于表2。

[0116] 表2

[0117] 实施例中使用的质粒

[0118]

质粒	描述	参考文献
pMLBAD	克隆载体, Tmp ^R	Lefebvre & Valvano (2002)
pMLBAD:Z3206 (SEQ ID NO: 23)	Z3206 在 pMLBAD 中, Tmp ^R , 表达受到阿拉伯糖诱导型启动子的控制	本文中的实施例
pMLBAD:gne (SEQ ID NO: 24)	gne 在 pMLBAD 中, Tmp ^R , 表达受到阿拉伯糖诱导型启动子的控制	本文中的实施例
pACYCpgl	空肠弯曲杆菌 <i>pgl</i> 簇 Cm ^R	Wacker 等人(2002)
pACYCgne::kan	在 <i>gne</i> 中含有 kan 盒的空肠弯曲杆菌 <i>pgl</i> 簇, Cm ^R , Kan ^R	Linton 等人(2005)
pWA2	可溶性周质6个His标记的AcrA, 在pBR322中处于Tet启动子的控制之下, Amp ^R	Feldman 等人(2005)

[0119] 材料-[1,6-³H]GlcNAc(30Ci/mmol)、UDP-[1-³H]GlcNAc(20Ci/mmol)和UDP-[6-³H]GalNAc(20Ci/mmol)得自American Radiolabeled Chemicals公司(St.Louis, MO)。Quantum 1硅胶G薄层板是Quantum Industries公司(Fairfield, NJ)的产品,而Baker Si250硅胶G板由Mallinckrodt Chemical Works公司制造。酵母膏和细菌用蛋白胨是BD Biosciences公司的产品。所有其它化学品都从标准商业来源获取。甲氧苄啶(50 μ g/ml)、氯霉素(20 μ g/ml)、氨苄青霉素(100 μ g/ml)和卡那霉素(50 μ g/ml)都按需要加到培养基中。

[0120] 重组质粒的构建—使用大肠杆菌菌株DH5 α 进行DNA克隆实验,所构建的质粒通过DNA测序进行验证。用寡核苷酸Z3206-Fw和Z3206-RvHA(AAACCCGGGATGAACGATAACG

TTTGTCTC(SEQ ID NO:17)和AAATCTAGATTAAGCGTAATCTGGAACATCGTATGGGTACTCAGAAACAAACGTTATGTC(SEQ ID NO:18);限制位点加上下画线),通过PCR从大肠杆菌0157:H45扩增Z3206基因。PCR片段用SmaI和XbaI消化后,连接到经SmaI-XbaI切割的pMLBAD载体中(Lefebvre, M.D.和Valvano M.A.(2002)Appl Environ Microbiol 68:5956-5964)。这导致得到质粒pMLBAD:Z3206(SEQ ID NO:23),其编码具有C-端血凝素标签的Z3206。

[0121] 用寡核苷酸gne-Fw和gne-RV(AAACCATGGATGAAAATTCTTATTAGCGG(SEQ ID NO:19)和AAATCTAGATTAAGCGTAATCTGGAACATCGTATGGGTAGCACTGTTTTTCCAATC(SEQ ID NO:20);限制位点加上下画线),从pACYCpgl扩增gne基因(Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M.(2002)Science 298,1790-1793),编码空肠弯曲杆菌pgl簇。PCR产物用NcoI和XbaI消化后,连接到pMLBAD的相同位点,得到质粒pMLBAD:gne(SEQ ID NO:24),其编码具有C-端血凝素标签的Gne(表2)。

[0122] 生长条件、蛋白质表达和免疫检测—大肠杆菌菌株在Luria-Bertani培养基(1%酵母膏,2%细菌用蛋白胨,0.6%NaCl)中于37℃进行培养并剧烈摇动。通过向生长至A₆₀₀为0.05-0.4的大肠杆菌细胞中加入终浓度为0.02-0.2%(w/v)的阿拉伯糖实现阿拉伯糖诱导的表达。诱导后5小时再次加入相同数量的阿拉伯糖,并继续培养4-15小时。

[0123] 分析程序—根据Pierce公司的生物技术最新报告“Eliminate Interfering Substances from Samples for BCA Protein Assay(消除BCA蛋白质测定中来自样品的干扰物质)”,在用脱氧胆酸和三氯乙酸使膜蛋白沉淀之后,利用BCA蛋白质测定法(Pierce)测出蛋白质浓度。在加入0.5ml 1%SDS和4ml Econosafe Economical Biodegradable Counting混合物(Research Products International, Corp., Mount Prospect, IL)之后,在Packard Tri-Carb 2100TR液体闪烁分光计中,通过闪烁光谱测定法分析样品的放射性。

[0124] 实施例1:编码GlcNAc-P-P-Und 4-差向异构酶的大肠杆菌0157基因的鉴定

[0125] 我们在本文中描述了一种新生物合成途径的惊奇发现,其中,通过4-差向异构酶先前未知的作用催化的GlcNAc-P-P-Und的4-OH差向异构化形成了GalNAc-P-P-Und。在这个途径中,GlcNAc-P-P-Und如下形成:从UDP-GlcNAc转移GlcNAc-P,由WecA催化,然后通过GlcNAc-P-P-Und-4-差向异构酶将GlcNAc-P-P-Und差向异构化成为GalNAc-P-P-Und,这是一个先前未知的途径(图2)。

[0126] 通过DNA同源性检索鉴定出GlcNAc-P-P-Und 4-差向异构酶候选物的编码基因。利用在<http://blast.ncbi.nlm.nih.gov/Blast.cgi>上找到的美国国家医学数据库的文库(the U.S.National Library of Medicine databases)进行同源性检索。筛选出编码在还原末端具有GalNAc的O抗原重复单位的不同细菌的基因组序列。将一个在还原末端含有GalNAc的重复单位的组和第二个在重复单位中缺乏末端GalNAc的组进行比较以鉴定潜在的差向异构酶。使用这些标准,Z3206鉴定为候选GlcNAc-P-P-Und 4-差向异构酶(表1)。

[0127] 在具有含GalNAc的O-抗原重复单位的大肠杆菌菌株中存在的GlcNAc 4-差向异构酶基因可以分成两个同源性组,如表1所示。令人惊奇地发现一个同源性组(含有gne1)明显地与GalNAc的存在相关联,因为O-抗原重复单位上的起始糖。进一步令人惊奇地发现第二组(含有gne2)表现出与UDP-Glc差向异构酶GalE高度的相似性,并存在于不起始O-抗原重复单位合成并具有GalNAc的大肠杆菌菌株中。大肠杆菌0157中的Z3206,一种与gne1具有高

度同源性的基因,鉴定为候选GlcNAc-P-P-Und 4-差向异构酶。Z3206基因的基因组位置与该途径中的作用相一致,因为它位于O-抗原簇的galF和属于荚膜异多糖酸簇的wcaM之间。

[0128] 实施例2-11描述的研究进一步证实了以上的发现,包括将GlcNAc 4-差向异构酶(大肠杆菌0157 Z3206)鉴定为催化GalNAc-P-P-Und的形成。

[0129] 实施例2:UDP-GalNAc不是大肠杆菌WecA(GlcNAc-磷酸转移酶)的底物

[0130] 为了确定大肠杆菌WecA将是否利用UDP-GalNAc作为GalNAc-P供体来形成GalNAc-P-P-Und,将来自大肠杆菌菌株K12、PR4019(一种WecA-过量表达菌株)和0157的膜部分与UDP-[³H]GalNAc一起孵育,该膜部分能合成据推测由GalNAc-P-P-Und的合成起始的在还原末端具有GalNAc的四糖O-抗原重复单位。

[0131] 大肠杆菌膜的制备—通过以1,000Xg离心10分钟收集细菌细胞,依次在冰冷的磷酸缓冲盐水中洗涤一次,用冷水洗涤一次,用10mM Tris-HCl(pH 7.4)、0.25M蔗糖洗涤一次。将细胞重新悬浮至在10mM Tris-HCl(pH 7.4)、0.25M蔗糖、10mM EDTA(含有0.2mg/ml溶菌酶)中密度为~200 A₆₀₀单位/ml,再于30℃孵育30分钟。通过以1,000Xg离心10分钟回收细菌细胞,快速重新悬浮在40体积的冰冷的10mM Tris-HCl(pH 7.4)中,再放置在冰上。10分钟后,用紧扣的杜恩斯匀浆器(tight-fitting Dounce homogenizer)将细胞匀浆15个冲程后,补充0.1mM苯甲基磺酰基氟和蔗糖至终浓度为0.25M。通过以1,000Xg离心10分钟去除未破碎的细胞,通过以40,000Xg离心20分钟回收细胞被膜。将膜部分重新悬浮于10mM Tris-HCl(pH 7.4)、0.25M蔗糖、1mM EDTA中,再次以40,000Xg沉降后,重新悬浮于相同缓冲液至蛋白质浓度为~20mg/ml。将膜部分保存在-20℃备用。

[0132] 在大肠杆菌膜中[³H]GlcNAc-P-P-Und和[³H]GalNAc-P-P-Und生物合成的体外测定—用于合成GlcNAc-P-P-Und和GalNAc-P-P-Und的反应混合物含有50mM Tris-HCl(pH 8)、40mM MgCl₂、5mM二硫苏糖醇、5mM 5' AMP、大肠杆菌膜部分(50-200μg膜蛋白和任一个5μM UDP-[³H]GlcNAc/GalNAc(500-2500dpm/pmol),总体积为0.05ml。于37℃孵育后,加入40体积的CHCl₃/CH₃OH(2:1)终止反应,按先前所述方法(Waechter, C. J., Kennedy, J. L. 和 Harford, J. B. (1976) Arch. Biochem. Biophys. 174, 726-737),制备含有[³H]HexNAc-P-P-十一异戊烯醇的总脂质提取物。在分配之后,将有机相在氮气流下干燥,再重新溶解于1ml CHCl₃/CH₃OH(2:1),取出等分试样(0.2ml),在闪烁小瓶中干燥后,在Packard Tri-Carb 2100TR液体闪烁分光计中通过液体闪烁光谱测定法分析其放射性。为了测定[³H]GlcNAc-P-P-Und或[³H]GalNAc-P-P-Und的合成速率,将脂质提取物在氮气流下干燥,再重新溶解于小体积的CHCl₃/CH₃OH(2:1)中,点样到10X 20cm硼酸盐浸渍的Baker Si250硅胶板上,该板用CHCl₃, CH₃OH, H₂O, 0.2M硼酸钠(65:25:2:2)展层。各个糖脂用Bioscan AR2000成像扫描仪(Bioscan, Washington, D.C.)进行检测。把[³H]GlcNAc/GalNAc-P-P-Und中的放射性总量乘以各个[³H]糖脂的百分率,计算出各糖脂的生物合成速率。

[0133] 将来自不同大肠杆菌菌株(K12、PR4019和0157)的膜部分与或者UDP-[³H]GlcNAc或者UDP-[³H]GalNAc一起孵育,如上所述测定[³H]GlcNAc/GalNAc-P-P-Und中的掺入。如表3所示,在与UDP-[³H]GalNAc一起孵育后未检测出标记的糖脂,当膜部分与UDP-[³H]GlcNAc一起孵育时仅GlcNAc-P-P-Und是可检测的。

[0134] 表3

[0135] 利用UDP-[³H]GlcNAc或者UDP-[³H]GalNAc作为底物在大肠杆菌膜部分中[³H]

GlcNAc/GalNAc-P-P-十一异戊烯醇的合成

[0136]

膜的来源	添加的糖核苷酸	形成的 $[^3\text{H}]$ 糖脂	
		GlcNAc-P-P-Und (<i>pmol/mg</i>)	GalNAc-P-P-Und (<i>pmol/mg</i>)
K12	UDP- $[^3\text{H}]$ GlcNAc	6.4	<0.01
K12	UDP- $[^3\text{H}]$ GalNAc	<0.01	<0.01
PR4019	UDP- $[^3\text{H}]$ GlcNAc	44	<0.01
PR4019	UDP- $[^3\text{H}]$ GalNAc	<0.01	<0.01
O157	UDP- $[^3\text{H}]$ GlcNAc	1.5	0.5
O157	UDP- $[^3\text{H}]$ GalNAc	<0.01	<0.01

[0137] 此外,加入外源Und-P后再与来自PR4019(WecA-过量表达菌株)的膜一起孵育和加入来自O157细胞的细胞溶质部分两者都没有导致由UDP-GalNAc形成GalNAc-P-P-Und。这些结果证明UDP-GalNAc不是WecA的底物,提示GalNAc-P-P-Und是由替代机制形成。

[0138] 当来自菌株K12的膜与UDP- $[^3\text{H}]$ GlcNAc一起孵育时,按预期合成了 $[^3\text{H}]$ GlcNAc-P-P-Und(Rush, J.S., Rick, P.D. 和 Waechter, C.J. (1997) Glycobiology, 7, 315-322)。然而,当来自菌株O157的膜与UDP- $[^3\text{H}]$ GlcNAc一起孵育时,除了 $[^3\text{H}]$ GlcNAc-P-P-Und之外,还观察到显示为 $[^3\text{H}]$ GalNAc-P-P-Und的第二种标记脂质(参见下文)。当检查两种糖脂的形成时程时, $[^3\text{H}]$ GlcNAc-P-P-Und(图1, ○)中放射性的掺入比 $[^3\text{H}]$ GalNAc-P-P-Und(图1, ●)中放射性的掺入发生更快,程度更高,这与前体-产物关系相匹配(图2)。

[0139] 大肠杆菌O157膜不利用UDP-GalNAc作为GalNAc-P供体进行GalNAc-P-P-Und合成的观察结果是一个实例,它证实了图2中说明的GalNAc-P-P-Und形成的生物合成途径。在该方案中,GlcNAc-P-P-Und如下形成:从UDP-GlcNAc转移GlcNAc-P,由WecA催化,然后通过先前未知的4-差向异构酶的作用使GlcNAc-P-P-Und差向异构化产生GalNAc-P-P-Und。

[0140] 实施例3:用来自大肠杆菌菌株O157的膜部分在体外形成的 $[^3\text{H}]$ GalNAc-P-P-Und的表征

[0141] 与图1检测到的额外O157-特异性糖脂产物相一致,作为GalNAc-P-P-Und,它对温和的碱甲醇解(甲苯/甲醇1:3,含有0.1N KOH, 0°C, 60min)是稳定的,通过在CHCl₃/CH₃OH/H₂O(10:10:3)中平衡的DEAE-纤维素保留,再用含有20mM乙酸铵的CHCl₃/CH₃OH/H₂O(10:10:3)洗脱,按先前关于 $[^3\text{H}]$ GlcNAc₁₋₂-P-P-Dol报道的(Waechter, C.J. 和 Harford, J.B. (1977) Arch. Biochem. Biophys. 181, 185-198)。

[0142] $[^3\text{H}]$ GalNAc-P-P-Und在硼酸盐浸渍的硅胶G上通过薄层层析法从 $[^3\text{H}]$ GlcNAc-P-P-Und中清楚地解析出来(Kean, E.L. (1966) J. Lipid Res. 7, 449-452)并通过制备型TLC纯化,如图3A和图3B所示。

[0143] 硼酸盐浸渍的薄层板和瓦特曼1号纸的制备—硅胶薄层板用硼酸钠浸透,即按照Kean的所述方法(Kean, E.L. (1966) J. Lipid Res. 7, 449-452),将硅胶薄层板短暂浸入2.5% Na₂B₄O₇ · 10H₂O的95%甲醇中。将该硼酸盐浸渍的TLC板在室温下干燥过夜后,放在Drierite上的真空干燥器内保存备用。刚好在层析之前,将板在短时间(~10-15min)加热至100°C而把板活化。将20X 30cm的瓦特曼1号(Whatman1)纸片浸渍在0.2M Na₂B₄O₇ · 10H₂O中而把瓦

特曼1号纸用硼酸钠浸渍。按照Cardini和Leloire的所述方法(Cardini, C.E.和Leloire, L.F. (1957) J. Biol. Chem. 225, 317-324), 把Whatman No.1纸片紧紧地压在两块Whatman No.3MM纸片之间并让其在室温下干燥数天。

[0144] 在体外反应中形成的聚糖产物的表征一个体糖脂($[^3\text{H}]\text{GalNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GlcNAc-P-P-Und}$)的聚糖在温和的酸水解释放之后通过下行纸层析法来表征。将GlcNAc/GalNAc脂质在氮气流下在带螺旋帽的锥形管中干燥并在0.2ml 0.01M HCl中加热到100℃达15分钟。水解后, 将样品加到含有0.4ml的AG50WX8(H^+)和0.4ml AG1X8(乙酸盐形式)的0.8ml混合床离子交换柱中, 用1.5ml水洗脱。洗脱液在氮气流下干燥, 重新溶解于小体积的 H_2O (0.02ml)中, 点样在一条长30cm的硼酸盐浸渍的瓦特曼1号纸条上, 用丁醇/吡啶/水(6:4:3)按下行模式展开40-50小时。干燥后, 把纸条切割成1cm的小长带, 用闪烁光谱测定法分析放射性。GlcNAc和GalNAc标准品用苯胺-二苯胺蘸浸(dip)试剂检测(Schwimmer, S. 和 Benvenue, A. (1956) Science 123, 543-544)。

[0145] 按照上述方法, 在温和的酸水解之后, 聚糖产物用0.1M NaBH_4 的0.1M NaOH(终体积0.1ml)还原而转变成它们的相应糖醇。在室温下温育过夜后, 反应物用几滴冰醋酸猝灭, 在氮气流下干燥使甲醇不含1滴醋酸, 重复几次。将糖醇溶于水, 过AG50WX8(H^+)和AG1X8(乙酸盐)的0.5ml柱脱盐, 在氮气下干燥, 点样到长30cm的Whatman No.3MM纸条上。Whatman No.3MM纸条用乙酸乙酯、吡啶、0.1M硼酸(65:25:20)按下行模式展开过夜, 干燥, 切割成1cm的小长带, 用闪烁光谱测定法分析放射性。GlcNAcitol和GalNAcitol标准品用改进的高碘酸盐-联苯胺(periodate-benzidine)蘸浸(dip)程序显现(Gordon, H.T., Thornburg, W. 和 Werum, L.N. (1956) Anal. Chem. 28, 849-855)。将纸条浸渍在丙酮、0.1M NaIO_4 (95:5)中, 让其风干3分钟, 然后浸渍在丙酮/乙酸/ H_2O /o-联甲苯胺(96:0.6:4.4:0.2gm)中, 含顺式二醇的糖醇在蓝色背景上染色成黄色斑点。

[0146] 糖脂的质谱测定法(“MS”)—用带有ABI Turbo V电喷雾离子源的ABI/MDS Sciex 4000Q-Trap混合式三节四极线性离子阱质谱仪(ABI/MDS-Sciex, Toronto, Canada)分析纯化的糖脂。简而言之, 将样品按10 $\mu\text{l}/\text{min}$ 灌到凭经验确定的离子源装置中, 通过将分子离子以线性离子阱模式碎裂获得了MS/MS(呈二维的质谱)信息。

[0147] 当糖脂用温和的酸(0.01N HCl, 100℃, 15min)处理时, 水溶性产物与 $[^3\text{H}]\text{GalNAc}$ 一起在下行纸层析上用硼酸盐浸渍的瓦特曼1号纸进行共层析分析(图3C)。另外, 当标记的糖还原时, 它被转变成 $[^3\text{H}]$ 糖醇GalNAc-OH(图3D)。此外, 负离子MS分析得出 $[\text{M-H}]^-$ 离子为 $m/z=1128$, 对于GalNAc-P-P-Und的预期, MS/MS子离子谱显示突出离子在 $m/z=907$, 对于含P-P-Und的糖脂的预期(Guan, Z., Breazeale, S.D. 和 Raetz, C.R. (2005) Anal. Biochem. 345, 336-339)。由菌株0157形成的糖脂产物鉴定为GalNAc-P-P-Und也得到了其形成来自外源GlcNAc-P-P-Und的支持(参见实施例7)。

[0148] 实施例4: 在表达Z3206基因的大肠杆菌细胞中用 $[^3\text{H}]\text{GlcNAc}$ 代谢性标记 $[^3\text{H}]\text{GalNAc-P-P-Und}$ (体内)

[0149] 为了研究大肠杆菌0157 Z3206基因的表达是否能够使细胞合成GalNAc-P-P-Und, 表达Z3206基因的大肠杆菌菌株21546(Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. 和 Rick, P.D. (1990) J. Biol. Chem., 265, 13490-13497)用 $[^3\text{H}]\text{GlcNAc}$ 进行代谢性标记并分析 $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$ 的形成。

[0150] 细菌细胞的代谢性标记—大肠杆菌细胞在剧烈摇动下在Luria-Bertani培养基中于37℃培养至A₆₀₀为0.5-1。加入³H]GlcNAc至终浓度为1μCi/ml,继续于37℃温育5分钟。加入0.5gm/ml碎冰终止放射性标记掺入到糖脂中,将培养物充分混合。以4000Xg离心10分钟来回收细菌细胞,弃去上清液。细胞用冰冷的磷酸缓冲盐水洗涤两次,通过剧烈涡旋混合而重新悬浮在10体积(细胞沉淀物)的甲醇中,用探测超声器按40%全功率短暂超声处理。超声处理后,加入20体积的氯仿,将提取物剧烈混合后,让其静置在室温下15分钟。不溶性物质通过离心沉淀出来,沉淀用小体积的CHCl₃/CH₃OH(2:1)反萃取两次。合并的有机提取物如下所述进行加工处理。

[0151] GlcNAc-P-P-Und和GalNAc-P-P-Und的纯化—GlcNAc/GalNAc-P-P-Und用CHCl₃/CH₃OH(2:1)抽提,按文献中描述的方法进行分配(Waechter, C. J., Kennedy, J. L. 和 Harford, J. B. (1976) Arch. Biochem. Biophys. 174, 726-737)使其没有水溶性物质。然后将有机提取物在氮气流下干燥,通过在含有0.1N KOH的甲苯/甲醇(1:3)于0℃脱酰基达60分钟使甘油磷酸脂质破坏。脱酰化反应用乙酸中和,用4体积的CHCl₃/CH₃OH(2:1)稀释,用1/5体积的0.9%NaCl洗涤。有机(下层)相用1/3体积的CHCl₃、CH₃OH、0.9%NaCl(3:48:47)洗涤,弃去水相。有机相用足够的甲醇稀释以在有机相中容纳残留的水相,上样到用CHCl₃/CH₃OH(2:1)平衡的DEAE-纤维素柱(5ml)上。柱子用20柱体积的CHCl₃/CH₃OH/H₂O(10:10:3)洗涤,然后用含有20mM乙酸铵的CHCl₃/CH₃OH/H₂O(10:10:3)洗脱。在硼酸盐浸渍的硅胶板上通过薄层层析解析(如较早所述)后,收集馏份(2ml),用茴香醛喷雾试剂监测放射性或GlcNAc/GalNAc-P-P-Und(Dunphy, P. J., Kerr, J. D., Pennock, J. F., Whittle, K. J. 和 Feeney, J. (1967) Biochim. Biophys. Acta 136, 136-147)。

[0152] 选择大肠杆菌菌株21546作为Z3206表达研究的宿主,因为UDP-ManNAcA合成中的突变导致利用GlcNAc-P-P-Und合成肠细菌共同抗原时出现阻断。因为大肠杆菌21546来源于大肠杆菌K12,所以它也不合成O-抗原重复片段(Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. 和 Reeves, P. (1994) J. Bacteriol., 176, 4144-4156),因此,大量的GlcNAc-P-P-Und累积用于转化成GalNAc-P-P-Und。当表达Z3206基因的菌株21546和转化体用³H]GlcNAc标记并且放射性标记脂质用薄层层析在硼酸盐浸渍的硅胶板上分析时,母菌株(图4A)仅合成一种标记的脂质GlcNAc-P-P-Und。然而,表达Z3206基因的21546细胞(图4B)也合成另外的标记脂质,显示为GalNAc-P-P-Und。

[0153] 实施例5:来自表达Z3206基因的大肠杆菌细胞的膜部分在体外合成GalNAc-P-P-Und

[0154] 为了证实由大肠杆菌0157 Z3206基因编码的蛋白质能催化GalNAc-P-P-Und的合成,将来自表达Z3206基因的大肠杆菌细胞的膜部分与³H]UDP-GlcNAc一起孵育,通过薄层层析(色谱制备和表征方法在实施例3中描述)在硼酸盐浸渍的硅胶板上分析³H]糖脂产物,如图5所示。当来自大肠杆菌K12或宿主菌株大肠杆菌21546细胞的膜部分与UDP-³H]GlcNAc一起孵育时,仅仅观察到³H]GlcNAc-P-P-Und(图5A和图5C)。然而,来自大肠杆菌0157和表达Z3206的大肠杆菌21546的膜部分也形成了GalNAc-P-P-Und(图5B和图5D)。

[0155] 实施例6:GlcNAc-P-P-Und但不是GalNAc-P-P-Und的形成在UMP存在下被逆转

[0156] 为了提供GalNAc-P-P-Und是从GlcNAc-P-P-Und而不是通过WecA的作用利用UDP-

GalNAc作为糖基供体合成的另外证据,检查了用UMP卸载内源预标记的 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的作用。被WecA催化的GlcNAc-磷酸转移酶反应通过加入过量的UMP而自由地逆转,即重新合成UDP-GlcNAc并释放Und-P。

[0157] 在这个实验当中,来自表达Z3206的大肠杆菌菌株21546的膜部分用UDP- $[^3\text{H}]\text{GlcNAc}$ 预先标记10分钟,然后加入1mM UMP,测定各标记的糖脂的剩余量。图6A所示结果表明在10分钟标记期结束时 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的相对量。在与1mM UMP一起孵育1分钟后,可以见到 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 的明显损失,而 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 峰相对来说无变化(图6B)(色谱制备和表征方法在实施例5中描述)。这一观察结果与表3中的结果相一致,说明WecA不催化从UDP-GalNAc将GalNAc-P转移到GalNAc-P-P-Und上。值得关注的是,在与UMP一起孵育的第2分钟(图6C),GlcNAc-P-P-Und的损失慢下来,并且在 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的峰中有略微减少,提示 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 通过差向异构酶反应的逆转用 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 库重新平衡(参见实施例7)。

[0158] 实施例7:由来自表达Z3206的大肠杆菌细胞的膜催化的外源纯化 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的相互转化

[0159] 为了提供GlcNAc-P-P-Und和GalNAc-P-P-Und可以通过来自表达Z3260的大肠杆菌细胞的膜部分直接相互转化的直接证据,将纯化的 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 作为外源底物进行了测试。

[0160] 纯化的 $[^3\text{H}]\text{GlcNAc-P-P-Und}/[^3\text{H}]\text{GalNAc-P-P-Und}$ 按实施例4中描述的方法(细菌细胞的代谢性标记和GlcNAc-P-P-Und和GalNAc-P-P-Und的纯化)来制备。按实施例2中关于在大肠杆菌膜中 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 生物合成的体外测定,将 $[^3\text{H}]\text{HexNAc-P-P-十一异戊烯醇}$ (2000dpm/pmol,在1%TritonX-100中分散,终浓度0.1%)与大肠杆菌膜一起孵育。

[0161] 初步实验表明当将外源 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 加入到在TritonX-100、CHAPS(3-[(3-胆酰胺丙基)二甲基铵基]-1-丙烷磺酸)、NonidetP-40或辛基葡糖苷中分散的反应混合物中时,差向异构酶是有活性的,并表现出pH最大在7-8.5的范围内。在与膜部分孵育之前纯化 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的色谱迁移率示于图7A和图7D。如图7B和图7E中见到的,与来自大肠杆菌21546的膜部分孵育不影响糖脂。然而,纯化糖脂与来自表达Z3206的大肠杆菌21546的膜部分的孵育能催化外源 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 向 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 的转变(图7C)和 $[^3\text{H}]\text{GalNAc-P-P-Und}$ 向 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 的转变(图7F)。这些结果直接证明GlcNAc-P-P-Und和GalNAc-P-P-Und在表达Z3206的大肠杆菌菌株中经酶催化而相互转变。

[0162] 实施例8:大肠杆菌Z3206不是UDP-GlcNAc 4-差向异构酶

[0163] 为了确定Z3206是否可以催化UDP-GalNAc的形成,使来自空肠弯曲杆菌的N-糖基化器(apparatus)在大肠杆菌中表达。在这个报道系统中,靶蛋白AcrA的糖基化依赖于pgl基因座的存在(Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wrenn, B. W., Aebi, M. (2002) Science 298, 1790-1793),包括功能性Gne UDP-Glc/UDP-GlcNAc差向异构酶(Bernatchez, S., Szymanski, C. M., Ishiyama, N., Li, J., Jarrell, H. C., Lau, P. C., Berghuis, A. M., Young, N. M., Wakarchuk, W. W. (2005) J. Biol. Chem. 280, 4792-4802)。如果pgl簇含有gne的缺失,那

么AcrA的糖基化就失去了(Linton,D.,Dorrell,N.,Hitchen,P.G.,Amber,S.,Karlyshev,A.V.,Morris,H.R.,Dell,A.,Valvano,M.A.,Aebi,M.和Wren,B.W.(2005)Mol Microbiol.55,1695-1703)。通过表达AcrA(pWA2)以及由Gne(pMLBAD:gne)或Z3206(pMLBAD:Z3206)代偿的pgl基因座 Δ gne研究了体外情况下在pgl操纵子 Δ gne存在下Z3206恢复AcrA-糖基化的能力。

[0164] 使用重新悬浮于100 μ l的上样缓冲液中浓度等于1 A₆₀₀单位的细胞,制备出总的大肠杆菌细胞提取物进行了免疫检测分析(Laemmli,U.(1970)Nature 227,680-685)。将10 μ l的等分试样上样到10%SDS-PAGE上。通过溶菌酶处理制备大肠杆菌细胞的周质提取物(Feldman,M.F.,Wacker,M.,Hernandez,M.,Hitchen,P.G.,Marolda,C.L.,Kowarik,M.,Morris,H.R.,Dell,A.,Valvano,M.A.,Aebi,M.(2005)Proc Natl Acad Sci USA 102,3016-3021),最终样品取10 μ l(相当于细胞的0.2 A₆₀₀单位)通过SDS-PAGE进行分析。在硝化纤维素膜上转印后,样品用特异性抗血清进行免疫染色(Aebi,M.,Gassenhuber,J.,Domdey,H.和te Heesen,S.(1996)Glycobiology 6,439-444)。使用抗AcrA(Wacker,M.,Linton,D.,Hitchen,P.G.,Nita-Lazar,M.,Haslam,S.M.,North,S.J.,Panico,M.,Morris,H.R.,Dell,A.,Wrenn,B.W.,Aebi,M.(2002)Science 298,1790-1793)抗体。使用抗兔IgG-HRP(Bio-Rad)作为第二抗体。用ECLTM蛋白质印迹分析检测试剂(Amersham Biosciences)进行检测。

[0165] 如图8所示,糖基化蛋白,它比非糖基化形式迁移慢,仅当表达pgl基因座 Δ gne的细胞被Gne代偿时才形成(第2泳道)。Z3206无法恢复报道糖蛋白的糖基化(图8,第1泳道)。因此,在依赖Gne的糖基化系统中,Z3206不会代偿AcrA的糖基化。Gne和膜结合Z3206的表达通过免疫检测予以证实。

[0166] 实施例9:弗氏志贺氏菌6+/-Z3206LPS的分析

[0167] 在图9显示了弗氏志贺氏菌6 O-抗原生物合成所需要的一些基因:编码核苷酸糖前体生物合成的酶的基因;编码糖基转移酶的基因;编码O抗原加工蛋白的基因;和编码负责O-乙酰化的蛋白的基因。O抗原的结构由Dmitriev,B.A等人进行了阐述(Dmitriev,B.A.等人,Somatic Antigens of Shigella(志贺氏菌属的菌体抗原)Eur J Biochem,1979.98:p.8;Liu B等人,Structure and genetics of Shigella O antigens(志贺氏菌属O抗原的结构和遗传学)FEMS Microbiology Review,2008.32:p.27)。

[0168] 为了鉴定弗氏志贺氏菌6 O-抗原生物合成所需要的所有基因,构建了基因组文库。

[0169] 弗氏志贺氏菌6基因组DNA的克隆 根据来自细菌的DNA分离方案,使用Macherey-Nagel NucleoSpin[®]Tissue试剂盒,分离出弗氏志贺氏菌6基因组DNA。从5个弗氏志贺氏菌6过夜培养物(各2ml)中分离出DNA,用100 μ l洗脱缓冲液(5mM Tris/HCl,pH 8.5)进行最后的洗脱。根据CopyControlTM Fosmid文库产生试剂盒(EPICENTRE)给出的方案所述,将洗脱的馏份合并,用异丙醇沉淀,将最终的沉淀物重新悬浮于52 μ l TE缓冲液,其总体积经历了终点修复。根据试剂盒方案所述,用1X TAE缓冲液,在1%低熔点琼脂糖凝胶上跑胶,来将终点修复的DNA纯化,回收后用乙醇沉淀。在7 μ l TE缓冲液中进行沉淀DNA的重新悬浮,再按照EPICENTRE方案,将其中的0.15 μ l DNA连接到pCC1FOS(SEQ ID NO:27)中。按照方案所述进行连接产物包装到噬菌体中。经包装的噬菌体在噬菌体稀释缓冲液中进行1:1稀释,取10 μ l

用来感染100 μ l EPI300-T1细胞,该细胞先前按EPICENTRE所述进行生长。该细胞(110 μ l)接种6块板,每板约100个菌落,使得6块板均含有完整弗氏志贺氏菌6基因组文库。板通过菌落转印来显现,阳性/阴性菌落进行蛋白质印迹分析和银染色。

[0170] 菌落转印法 对于菌落转印,把硝化纤维素膜放在固体琼脂板上面,再取下来,放入1X PBST中洗涤三次并按同样的方式进行处理。将膜首先在10%乳中在室温下封闭1小时,然后在室温下在2ml 1%乳(在PBST中)中与抗VI型抗血清(第一抗体)一起保温。在PBST中洗涤三次(每次10分钟)后,将膜与第二抗体、1:20000过氧化物酶缀合的山羊-抗兔IgG(BioRad)(在2ml 1%乳(在PBST中)中)在室温下再保温另外1小时。用PBST进行最后三次洗涤(每次10分钟)后,将膜在由SuperSignal®West Dura Extended Duration底物试剂盒(Thermo Scientific)提供的具有鲁米诺(luminol)和过氧化物缓冲液的1:1混合物的UVP Chemi Doc成像系统中显影。

[0171] 在产生弗氏志贺氏菌6基因组文库之后,与弗氏志贺氏菌6抗血清反应的克隆用Liu等人(Liu等人,2008)先前测序的区域的引物步查从rmlB到wfbZ进行测序(图9)。引物rmlB_rev和wfbZ_fwd(弗氏志贺氏菌-Z3206)在rmlB和wfbZ中退火并用来测定克隆插入片段的序列直至分别达到wcaM和hisI/F(弗氏志贺氏菌+Z3206)(图10)。

[0172] 为了建立O抗原合成是否在缺乏Z3206(因此阻碍und-GlcNAc向und-GalNAc的差向异构化)的克隆中保持,构建了两种质粒(SEQ ID NO.28和SEQ ID NO.29)(图10),转化到大肠杆菌细胞中并通过银染色和蛋白质印迹进行分析。

[0173] 如图11所示,在大肠杆菌细胞+或-Z3206中产生了LPS。O抗原在没有Z3206时可以产生,然而产量较低,这表明没有差向异构酶(Z3206)的多糖产生效率较低。

[0174] 实施例10:弗氏志贺氏菌6+/-Z3206LL0的分析

[0175] 通过C18柱层析纯化十一异戊烯醇-PP-O抗原 使表达弗氏志贺氏菌(*S.flexeneri*)抗原+/-Z3206的大肠杆菌细胞沉淀,在50ml 0.9%NaCl中洗涤一次,将终沉淀物冻干过夜。该沉淀物在30ml 85-95%甲醇中洗涤一次,用10:10:3氯仿-甲醇-水(v/v/v)反萃取,通过加水将提取物转变成两相Bligh/Dyer系统,导致最终比率为10:10:9(C:M:W)。通过离心分离各相,上层水相分别上样到用10ml甲醇调节并用10ml 3:48:47(C:M:W)平衡的C18 Sep-Pak柱体上。上样后,柱体用10ml 3:48:47(C:M:W)洗涤并用5ml 10:10:3(C:M:W)洗脱。将C18柱的上样、流通、洗涤和洗脱的20个OD样品在Eppendorf Concentrator Plus管中干燥,用250 μ l甲醇洗涤,重蒸发后用30 μ l ddH₂O多洗涤三次。

[0176] 糖脂水解。来自C18柱洗涤的糖脂样品如下水解:将干燥的样品溶解在2ml正丙醇:2M三氟乙酸(1:1)中,加热到50℃达15分钟并在N₂下蒸发至干。

[0177] 用氨基苯甲酸盐和HPLC进行寡糖标记。按照Bigge等人(Bigge,1995)中描述的方法进行标记,按照Merry等人(2002)(Merry等人,2002)所述的纸盘方法进行聚糖提纯(cleanup)。按照Royle等人(Royle,2002)中描述的方法但修改到三个溶剂系统,采用GlycoSep-N正相柱,通过HPLC进行2-AB标记的聚糖的分离。溶剂A是10mM甲酸铵pH 4.4的80%乙腈。溶剂B是30mM甲酸铵pH 4.4的40%乙腈。溶剂C是0.5%甲酸。柱温为30℃,并通过荧光(λ_{ex} =330nm, λ_{em} =420nm)检测2-AB标记的聚糖。梯度条件是100%A至100%B的线性梯度,历经160分钟,流速为0.4ml/min,然后2分钟100%B至100%C,流速至1ml/min。柱用100%C洗涤5分钟,回到100%A历经2分钟后,用100%A按流速为1ml/min流过15分钟,然后回到流速至

0.4ml/min达5分钟。所有样品都在水中注射。

[0178] 将表达有(SEQ ID NO:29)或无(SEQ ID NO:28)Z3206的弗氏志贺氏菌O抗原的质粒转化到SCM3细胞中(图10)。迟洗脱体积的痕迹显示含有弗氏志贺氏菌O抗原+/-Z3206的两个样品曲线间的差异(图12)。这种在洗脱模式中的差异可以用在还原端带有不同单糖的不同寡糖结构进行解释:GlcNAc或GalNAc取决于差向异构酶(Z3206)的存在。

[0179] 实施例11:通过由弗氏志贺氏菌6+/-Z3206产生的生物缀合物的制备和表征分析pglB的特异性

[0180] 为了评价PglB是否可以将在还原端具有GlcNAc(弗氏志贺氏菌60-抗原)的寡糖转移到载体蛋白EPA Nickel上,使用抗EPA和抗VI型抗体,通过蛋白质印迹法分析来自表达EPA(SEQ ID NO:25)、PglB(SEQ ID NO:26)和弗氏志贺氏菌60-抗原+/-Z3206(SEQ ID NO:29/SEQ ID NO:28)的大肠杆菌细胞的纯化提取物。将在还原端有和无GalNAc的弗氏志贺氏菌O6抗原通过PglB转移到EPA上并用抗EPA和抗VI抗血清进行检测(图13)。

[0181] O抗原仍可产生和检出,但产量较低,这表明没有差向异构酶的多糖产生效率较低。

[0182] 尽管本发明参照其实施方案进行了具体地说明和描述,但是本领域技术人员应当理解,在不偏离由权利要求书涵盖的发明范围的情况下,可以在形式和细节上对本发明进行各种变化。本领域技术人员将理解为落入本发明的范围之内的这些各种变化,特别是包括含有除来自在还原末端具有GalNAc的大肠杆菌和弗氏志贺氏菌以外的聚糖的N-糖基化蛋白和生物缀合物。

[0183] 序列表

[0184] 申请人:GlycoVaxyn AG

[0185] 名称:在原核细胞中产生免疫原性多糖的生物合成系统

[0186] SEQ ID NO的数目:29

[0187] SEQ ID NO:1-大肠杆菌0157 Z3206的核苷酸序列

[0188] 长度:993

[0189] 类型:DNA

[0190] 生物体:大肠杆菌0157

[0191] 序列:

[0192] ATGAACGATAACGTTTTGCTCATAGGAGCTTCCGGATTTCGTAGGAACCCGACTACTT

[0193] GAAACGGCAATTGCTGACTTTAATATCAAGAACCTGGACAAACAGCAGAGCCACTTT

[0194] TATCCAGAAATCACACAGATTGGCGATGTTTCGCGATCAACAGGCACTCGACCAGGCG

[0195] TTAGTCGGTTTTGACACTGTTGTACTACTGGCAGCGGAACACCGCGATGACGTCAGC

[0196] CCTACTTCTCTCTATTATGATGTCAACGTTTCAGGGTACCCGCAATGTGCTGGCGGCC

[0197] ATGGAAAAAATGGCGTTAAAAATATCATCTTTACCAGTTCCGTTGCTGTTTATGGT

[0198] TTGAACAAACACAACCCTGACGAAAACCATCCACACGACCCTTTCAACCACTACGGC

[0199] AAAAGTAAGTGGCAGGCAGAGGAAGTGCTGCGTGAATGGTATAACAAAGCACCAACA

[0200] GAACGTTTATTAACCATCATCCGTCCTACCGTTATCTTCGGTGAACGCAACCGCGGT

[0201] AACGTCTATAACTTGCTGAAACAGATCGCTGGCGGCAAGTTTATGATGGTGGGCGCA

[0202] GGGACTAACTATAAGTCCATGGCTTATGTTGGAACATTGTTGAGTTTATCAAGTAC

- [0203] AAAGTGAAGAATGTTGCCGAGGTTATGAGGTTTATAACTACGTTGATAAGCCAGAC
[0204] CTGAACATGAACCAGTTGGTTGCTGAAGTTGAACAAAGCCTGAACAAAAAGATCCCT
[0205] TCTATGCACTTGCCTTACCCACTAGGAATGCTGGGTGGATATTGCTTTGATATCCTG
[0206] AGCAAAATTACGGGCAAAAAATACGCTGTCTAGCTCAGTGCAGTGAAGAAAAATCTGC
[0207] GCAACAACACAGTTTGACGCAACGAAAGTGCATTCTTCAGGTTTTGTGGCACCGTAT
[0208] ACGCTGTCTGCAAGGTCTGGATCGAACACTGCAGTATGAATTCGTTTCATGCCAAAAAA
[0209] GACGACATAACGTTTGTCTCTGAG
[0210] SEQ ID NO:2-Z3206的氨基酸序列
[0211] 长度:331
[0212] 类型:PRT
[0213] 生物体:大肠杆菌0157
[0214] 序列:
[0215]

MNDNVLLIGASGFVGTTRLLETAIADFNILNDKQQSHFYPEITQIGDVRDQQALDQALVGFDTVVLLAAEHRDDVSP
TSLYYDVNVQGTNRNLAAMEKNGVKNIIFTSSVAVYGLNKHNPDENHPDHPFNHYGKSKWQAEEVLREWYNKAPTER
SLTII RPTVIFGERNRGNVYNLLKQIAGGKFMVAGAGTNYKSMAYVGNIVEFIKYKLKNVAAGYEVYNYVDKPDLM
NQLVAEVEQSLNKKIPSMHLPYPLGMLGGYCFDILSKITGKKYAVSSVRVKKFCATTQFDATKVHSSGFVAPYTLSQ
GLDRTLQYEFVHAKKDDITFVSE

- [0216] SEQ ID NO:3-大肠杆菌055gne的核苷酸序列
[0217] 基因座 AF461121_1 BCT 02-MAY-2002
[0218] 定义(UDP-GlcNAc 4-差向异构酶Gne[大肠杆菌(*Escherichia coli*)])
[0219] 检索号AAL67550
[0220] 长度:993
[0221] 类型:DNA
[0222] 生物体:大肠杆菌055
[0223] 序列:

[0224] ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCTG TAGGAACCCG ACTACTTGAA
ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC AGCAGAGCCA CTTTTATCCA GAAATCACAC
AGATTGGTGA TGTTCTGTAT CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTGCTACT
GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG TCAACGTTCA GGGTACCCGC
AATGTGCTGG CGGCCATGGA AAAAAATGGC GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG
GTTTGAACAA ACACAACCCT GACGAAAACC ATCCACACGA TCCTTTCAAC CACTACGGCA AAAGTAAGTG
GCAGGCAGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA CCAACAGAAC GTTCATTAAC CATCATCCGT
CCTACCGTTA TCTTCGGTGA ACGGAACCGC GGTAACGTCT ATAAGTTGCT GAAACAGATC GCTGGCGGCA
AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT GGAAACATTG TTGAGTTTAT
CAAGTACAAA CTGAAGAATG TTGCCGAGG TTACGAGGTT TATAACTACG TTGATAAGCC AGACCTGAAC
ATGAACCAGT TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC TTGCCTTACC
CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG CAAAATTACG GGCAAAAAAT ACGCTGTCTG
CTCTGTGCGC GTGAAAAAAT TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT

GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA ATTCGTCCAT GCCAAAAAAG
ACGACATAAC GTTTGTCTTCT GAG

[0225] SEQ ID NO:4-大肠杆菌055UDP-GlcNAc 4-差向异构酶Gne的氨基酸序列

[0226] 基因座 AF461121_1

[0227] 定义(UDP-GlcNAc 4-差向异构酶Gne[大肠杆菌])

[0228] 检索号AAL67550

[0229] 长度:331aa线性

[0230] 类型:PRT

[0231] 生物体:大肠杆菌055

[0232] 序列:

[0233] mndnvlliga sgfvgrllle taiadfnikn ldkqqshfyp eitqigdvrd

[0234] qqaldqalag fdtvllaae hrddvsptsl yydvnnvgtr nvlaamekng

[0235] vkniiftssv avyglmkhnp denhphdpfn hygkskwqae evlrewynka

[0236] ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv

[0237] gnivefikyk lknvaagyev ynyvdkpdlm mnqlvaeveq slnkkipsmh

[0238] lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf

[0239] vapytllsqgl drtlqyefvh akkdditfvs e

[0240] SEQ ID NO.5-大肠杆菌086gne1的核苷酸序列

[0241] 基因座 AA037706 BCT 06-DEC-2005

[0242] 定义UDP-GlcNAc C4-差向异构酶[大肠杆菌086]。

[0243] 检索号AA037706

[0244] 长度:993

[0245] 类型:DNA

[0246] 生物体:大肠杆菌086

[0247] 序列:

[0248] ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCT TAGGAACCCG ACTACTTGAA

[0249] ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC AGCAGAGCCA CTTTTATCCA

[0250] GAAATCACAC AGATTGGTGA TGTTCTGTAT CAACAGGCAC TCGACCAGGC GTTAGCCGGT

[0251] TTTGACACTG TTGTACTACT GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC

[0252] TATTATGATG TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC

[0253] GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA ACACAACCCT

[0254] GACGAAAACC ATCCACACGA CCCTTTCAAC CACTACGGCA AAAGCAAGTG GCAGGCGGAG

[0255] GAAGTGCTGC GTGAATGGTA TAACAAAGCA CCAACAGAAC GTTCATTAAC TATCATCCGT

[0256] CCTACCGTTA TCTTCGGTGA ACGCAACCGC GGTAACGTCT ATAAGTTGCT GAAACAGATC

[0257] GCTGGCGGCA AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT

[0258] GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGAGG TTACGAGGTT

[0259] TATAACTACG TTGATAAGCC AGACCTGAAC ATGAACCAGT TGGTTGCTGA AGTTGAACAA

[0260] AGCCTGAACA AAAAGATCCC TTCTATGCAC TTGCCTTACC CACTAGGAAT GCTGGGTGGA

[0261] TATTGCTTTG ATATCCTGAG CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC

- [0262] GTGAAAAAAT TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
- [0263] GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA ATTCGTCCAT
- [0264] GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG
- [0265] SEQ ID NO:6-大肠杆菌086UDP-GlcNAc C4-差向异构酶的氨基酸序列
- [0266] 基因座 AA037706
- [0267] 定义UDP-GlcNAc C4-差向异构酶[大肠杆菌086]。
- [0268] 检索号AA037706
- [0269] 长度:331aa线性
- [0270] 类型:PRT
- [0271] 生物体:大肠杆菌086
- [0272] 序列:
- [0273] mndnvlliga sgfvgrlle taiadfnikn ldkqqshfyp eitqigdvrd
- [0274] qqaldqalag fdtvllaae hrddvsptsl yydvnvqgtr nvlaamekng
- [0275] vkniiftssv avyglkhnp denhphdpfn hygkskwqae evlrewynka
- [0276] ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv
- [0277] gnivefikyk lknvaagyev ynyvdkpdlm mnqlvaeveq slnkkipsmh
- [0278] lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf
- [0279] vapytllsqgl drtlqyefvh akkdditfvs e
- [0280] SEQ ID NO:7-鲍氏志贺氏菌(*Shigella boydii*)018gne的核苷酸序列
- [0281] 基因座ACD09753 BCT 05-MAY-2008
- [0282] 定义UDP-N-乙酰基葡萄糖胺4-差向异构酶[鲍氏志贺氏菌(*Shigella boydii*) CDC3083-94]。
- [0283] 检索号ACD09753
- [0284] 长度:993
- [0285] 类型:DNA
- [0286] 生物体:鲍氏志贺氏菌(*Shigella boydii*)018
- [0287] 序列:
- [0288] ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCT TAGGAACCCG ACTACTTGAA
- [0289] ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC AGCAGAGCCA TTTTATCCA
- [0290] GCAATCACAC AGATTGGCGA TGTTCTGTAT CAACAGGCAC TCGACCAGGC GTTAGCCGGT
- [0291] TTTGACACTG TTGTACTACT GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC
- [0292] TATTATGATG TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC
- [0293] GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA ACACAACCCT
- [0294] GACGAAAACC ATCCACACGA CCCTTTCAAC CACTACGGCA AAAGTAAGTG GCAGGCAGAG
- [0295] GAAGTGCTGC GTGAATGGTA TAACAAAGCA CCAACAGAAC GTTCATTAAC CATCATCCGT
- [0296] CCTACCGTTA TCTTCGGTGA ACGCAACCGC GGTAACGTCT ATAAGTTGCT GAAACAGATC
- [0297] GCTGGCGGCA AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT
- [0298] GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGAGG TTATGAGGTT
- [0299] TATAACTATG TTGATAAGCC AGACCTGAAC ATGAACCAGT TGGTTGCTGA AGTTGAACAA

- [0300] AGCCTGAACA AAAAGATCCC TTCTATGCAC TTGCCTTACC CACTAGGAAT GCTGGGTGGA
- [0301] TATTGCTTTG ATATCCTGAG CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC
- [0302] GTGAAAAAAT TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
- [0303] GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA ATTCGTCCAT
- [0304] GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG
- [0305] SEQ ID NO:8-鲍氏志贺氏菌(*Shigella boydii*)018UDP-N-乙酰基葡萄糖胺4-差向异构酶的氨基酸序列
- [0306] 基因座ACD09753
- [0307] 定义UDP-N-乙酰基葡萄糖胺4-差向异构酶[鲍氏志贺氏菌(*Shigella boydii*) CDC3083-94]。
- [0308] 检索号ACD09753
- [0309] 长度:331aa线性
- [0310] 类型:PRT
- [0311] 生物体:鲍氏志贺氏菌(*Shigella boydii*)018
- [0312] 序列:
- [0313] mndnvlliga sgfvgrlle taiadfnikn ldqkqshfyp aitqigdvrdd qqaldqalag
- [0314] fdtvllaae hrddvsptsl yydvvnqgtr nvlaamekng vkniiftssv avyglknkhn
- [0315] denhphdpfn hygkskwqae evlrewynka ptersltiir ptvifgernr gnvynllkqi
- [0316] aggkfmmvga gtnyksmayv gnivefikyk lknvaagyev ynyvdkpdlm mnqlvaeveq
- [0317] slnkkipsmh lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf
- [0318] vapytllsqgl drtlqyefvh akkdditfvs e
- [0319] SEQ ID NO:9-肠沙门氏菌(*Salmonella enterica*)030gne的核苷酸序列
- [0320] 基因座 AAV34516 BCT 25-OCT-2004
- [0321] 定义UDP-GlcNAc 4-差向异构酶[肠沙门氏菌(*Salmonella enterica*)亚种 *salamae* serovar Greenside]。
- [0322] 检索号AAV34516
- [0323] 长度:993
- [0324] 类型:DNA
- [0325] 生物体:肠沙门氏菌(*Salmonella enterica*)030
- [0326] 序列:
- [0327] ATGAACGATA ACGTTTTGCT CATTGGTGCT TCCGGATTCT TAGGAACCCG ACTCCTTGAA
- [0328] ACGGCAGTGG ATGATTTTAA TATCAAGAAC CTGGATAAAC AGCAAAGCCA TTTCTACCCA
- [0329] GAGATTACAC ACATTGGCGA TGTTCTGTGAC CAACAAATCC TTGACCAGAC GTTGGTGGGT
- [0330] TTTGACACCG TAGTACTATT GGCTGCGGAG CATCGTGATG ATGTTAGTCC TACCTCGCTT
- [0331] TATTATGATG TCAACGTCCA GGGAACGCGT AATGTACTGG CGGCGATGGA AAAAAATGGT
- [0332] GTAAAAAATA TCATTTTAC CAGTTCCGTT GCAGTTTATG GACTCAACAA GAAAAATCCT
- [0333] GACGAAACGC ACCCTCACGA TCCCTTTAAT CATTACGGAA AAAGTAAATG GCAAGCAGAA
- [0334] GAAGTTCTGC GTGAGTGGCA TGCTAAAGCG CCGAATGAGC GTTCTTTGAC CATAATTCGT
- [0335] CCTACCGTTA TTTTCGGGGA GCGTAACCGC GGTAATGTAT ACAATCTCTT GAAACAGATC

- [0336] GCTGGTGGTA AATTTGCGAT GGTTGGTCCG GGAACAACT ATAAATCAAT GGCTTATGTT
- [0337] GGTAATATCG TTGAGTTTAT CAAATTCAAA CTCAAGAATG TTACGGCGGG CTATGAAGTT
- [0338] TATAATTATG TTGATAAACC TGATCTGAAT ATGAATCAAT TGGTTGCTGA AGTAGAGCAG
- [0339] AGCCTGGGCA AAAAAATACC ATCGATGCAC CTTCCATATC CATTAGGTAT GCTGGGGGGT
- [0340] TACTGTTTCG ATATCCTGAG CAAAGTAACG GGCAAGAAGT ACGCTGTAAG TTCGGTTCGT
- [0341] GTTAAAAAAT TCTGTGCGAC AACGCAGTTT GATGCAACAA AAGTGCATTC TTCTGGTTTT
- [0342] GTTGCGCCAT ACACCTTATC TCAGGGGTTG GATCGTACAC TGCAATATGA ATTTGTTTCAT
- [0343] GCAAAGAAAG ATGACATTAC ATTCGTTTCA GAG
- [0344] SEQ ID NO:10-肠沙门氏菌(*Salmonella enterica*)030UDP-GlcNAc 4-差向异构酶的氨基酸序列
- [0345] 基因座 AAV34516
- [0346] 定义UDP-GlcNAc 4-差向异构酶[肠沙门氏菌(*Salmonella enterica*)亚种 *salamaeserovar* Greenside]。
- [0347] 检索号AAV34516
- [0348] 长度:331aa线性
- [0349] 类型:PRT
- [0350] 生物体:肠沙门氏菌(*Salmonella enterica*)030
- [0351] 序列:
- [0352] mndnvllica sgfvgrlle tavddfnikn ldkqqshfyp eithigdvr d qqildqtlvg
- [0353] fdtvllaae hrddvsptsl yydvnvqgtr nvlaamekng vkniiftssv avyglmknp
- [0354] dethphdpfn hygkskwqae evlrewhaka pnersltiir ptvifgernr gnvynllkqi
- [0355] aggkfamvgp gtnyksmayv gnivefikfk lknvtagyev ynyvdkpdl n mnqlvaeveq
- [0356] slgkkipshm lpyplgmllg ycfdilskvt gkkyavssvr vkkfcattqf datkvhssgf
- [0357] vapytllsqgl drtlqyefvh akkdditfvs e
- [0358] SEQ ID NO:11-空肠弯曲杆菌(*C. jejuni*)gne的核苷酸序列
- [0359] 基因座 YP_002344524 BCT 14-SEP-2010
- [0360] 定义UDP-GlcNAc/Glc 4-差向异构酶[空肠弯曲杆菌空肠亚种(*Campylobacter jejuni* subsp. *jejuni*)
- [0361] 检索号 YP_002344524
- [0362] 长度:987
- [0363] 类型:DNA
- [0364] 生物体:空肠弯曲杆菌(*C. jejuni*)
- [0365] 序列:
- [0366] ATGAAAATTCTTATTAGCGGTGGTGCAGGTTATATAGGTTCTCATACTTTAAGACAAT
- [0367] TTTTAAAAACAGATCATGAAATTTGTGTTTTAGATAATCTTTCTAAGGGTTCTAAAAAT
- [0368] CGCAATAGAAGATTTGCAAAAAACAAGAGCTTTTAAATTTTTCGAACAAGATTTAAGT
- [0369] GATTTTCAAGGCGTAAAAGCATTGTTTGAGAGAGAAAAATTTGACGCTATTGTGCATT
- [0370] TTGCAGCAAGCATTGAAGTTTTTGAAAGTATGCAAAATCCTTTAAATATTATATGAA
- [0371] CAACACTGTTAATACGACAAATCTCATCGAACTTGTTTGCAAACTGGAGTGAATAAA

- [0372] TTTATATTTTCTTCAACGGCGGCCACTTATGGCGAACCACAAACTCCCGTTGTGAGCG
- [0373] AAACAAGTCCTTTAGCACCTATTAATCCTTATGGGCGTAGTAAGCTTATGAGTGAAGA
- [0374] AGTTTTGCGTGATGCAAGTATGGCAAATCCTGAATTTAAGCATTGTATTTTAAGATAT
- [0375] TTTAATGTTGCAGGTGCTTGTATGGATTATACTTTAGGACAACGCTATCCAAAAGCGA
- [0376] CTTTGCTTATAAAAGTTGCAGCTGAATGTGCCGAGGAAAACGTGATAAACTTTTCAT
- [0377] ATTTGGCGATGATTATGATACAAAAGATGGTACTTGCATAAGAGATTTTATCCATGTA
- [0378] GATGATATTTCAAGTGCACATTTAGCGGCTTTGGATTATTTAAAAGAGAATGAAAGCA
- [0379] ATGTTTTTAATGTAGGTTATGGACATGGTTTTAGCGTAAAAGAAGTGATTGAAGCGAT
- [0380] GAAAAAAGTTAGCGGAGTGGATTTTAAAGTAGAACTTGCCCCACGCCGTGCGGGTGAT
- [0381] CCTAGTGTATTGATTTCTGATGCAAGTAAATCAGAAATCTTACTTCTTGGCAGCCTA
- [0382] AATATGATGATTTAGAGCTTATTTGTAAATCTGCTTTTGATTGGGAAAAACAGTGTTA
- [0383] A
- [0384] SEQ ID NO:12-空肠弯曲杆菌(*C. jejuni*)UDP-GlcNAc/Glc 4-差向异构酶的氨基酸序列
- [0385] 基因座 YP_002344524
- [0386] 定义UDP-GlcNAc/Glc 4-差向异构酶[空肠弯曲杆菌空肠亚种(*Campylobacter jejuni* subsp. *jejuni*)
- [0387] 检索号 YP_002344524
- [0388] 长度:328aa线性
- [0389] 类型:PRT
- [0390] 生物体:空肠弯曲杆菌(*C. jejuni*)
- [0391] 序列:
- [0392] mkilisggag yigshtlrqf lktdheicvl dnlskgskia iedlqktraf
- [0393] kffeqdl sdf qgvkalfere kfdaivhfaa sievfesmqn plkyymmntv
- [0394] nttnlietcl qtgvnkfifs
- [0395] staatygepq tpvvsetspl apinpygrsk lmseevlrda smanpefkhc
- [0396] ilryfnvaga cmdytlgqry pkatllikva aecaagkrdk lfifgddydt
- [0397] kdgtcirdfi hvddissahl
- [0398] aaldylkene snvfnvgygh gfsvkeviea mkkvsgvdfk velaprragd
- [0399] psvlisdask irnltswqpk yddlelicks
- [0400] afdwekqc
- [0401] SEQ ID NO:13-大肠杆菌K12galE的核苷酸序列
- [0402] 基因座 AP_001390 BCT 30-APR-2010
- [0403] 定义UDP-半乳糖-4-差向异构酶[大肠杆菌菌株K-12亚菌株W3110]。
- [0404] 检索号 AP_001390
- [0405] 长度:1,017
- [0406] 类型:DNA
- [0407] 生物体:大肠杆菌K12
- [0408] 序列:

- [0409] ATGAGAGTTCTGGTTACCGGTGGTAGCGGTACATTGGAAGTCATACCTGTGTGCAAT
[0410] TACTGCAAAACGGTCATGATGTCATCATTCTTGATAACCTCTGTAACAGTAAGCGCAG
[0411] CGTACTGCCTGTTATCGAGCGTTTAGGCGGCAAACATCCAACGTTTGTGAAGGCGAT
[0412] ATTCGTAACGAAGCGTTGATGACCGAGATCCTGCACGATCACGCTATCGACACCGTGA
[0413] TCCACTTCGCCGGGCTGAAAGCCGTGGGCGAATCGGTACAAAAACCGCTGGAATATTA
[0414] CGACAACAATGTCAACGGCACTCTGCGCCTGATTAGCGCCATGCGCGCCGCTAACGTC
[0415] AAAAACTTTATTTTGTAGCTCCTCCGCCACCGTTTATGGCGATCAGCCCCAAAATTCAT
[0416] ACGTTGAAAGCTTCCCGACCGGCACACCGCAAAGCCCTTACGGCAAAAGCAAGCTGAT
[0417] GGTGGAACAGATCCTCACCGATCTGCAAAAAGCCCAGCCGGACTGGAGCATTGCCCTG
[0418] CTGCGCTACTTCAACCCGGTTGGCGCGCATCCGTCCGGCGATATGGGCGAAGATCCGC
[0419] AAGGCATTCCGAATAACCTGATGCCATACATCGCCCAGGTTGCTGTAGGCCGTCGCGA
[0420] CTCGCTGGCGATTTTTGGTAACGATTATCCGACCGAAGATGGTACTGGCGTACGCGAT
[0421] TACATCCACGTAATGGATCTGGCGGACGGTCACGTCGTGGCGATGGAAAACTGGCGA
[0422] ACAAGCCAGGCGTACACATCTACAACCTCGGCGCTGGCGTAGGCAACAGCGTGCTGGA
[0423] CGTGGTTAATGCCTTCAGCAAAGCCTGCGGCAAACCGGTTAATTATCATTTTGCACCG
[0424] CGTCGCGAGGGCGACCTTCCGGCCTACTGGGCGGACGCCAGCAAAGCCGACCGTGAAC
[0425] TGAAGTGGCGCGTAACGCGCACACTCGATGAAATGGCGCAGGACACCTGGCACTGGCA
[0426] GTCACGCCATCCACAGGGATATCCCGATTAA
[0427] SEQ ID NO:14-大肠杆菌K12UDP-半乳糖-4-差向异构酶的氨基酸序列
[0428] 基因座 AP_001390
[0429] 定义UDP-半乳糖-4-差向异构酶[大肠杆菌菌株K-12亚菌株W3110]。
[0430] 检索号 AP_001390
[0431] 长度:338aa线性
[0432] 类型:PRT
[0433] 生物体:大肠杆菌K12
[0434] 序列:
[0435] mrvltvgsgyigshtcvqllqngdhviildnlenskrsvlpvierlggkhptfvegd
[0436] irnealmteilhdhaidtvihfaglkavgesvqkpleyydnnvngtllrlisamraanv
[0437] knfifsssatvygdqpkipyvesfptgtpqspygksklmveqiltdlqkaqpdwsial
[0438] lryfnpvgahpsgdmgedpqgipnnlmpyiaqvavgrrdslaifgndyptedgtgvr
[0439] yihvmdladghvameklankpgvhiynlgagvgnsvl dvvnafskacgkpvnyhfap
[0440] rregdlpaywadaskadrelnwrvttrtldemaqdtwhwqsrhpqgypd
[0441] SEQ ID NO:15-大肠杆菌086gne2的核苷酸序列
[0442] 基因座 AAV85952 BCT 27-MAR-2005
[0443] 定义Gne[大肠杆菌086]。
[0444] 检索号AAV85952
[0445] 长度:1,020
[0446] 类型:DNA
[0447] 生物体:大肠杆菌086

[0448] 序列:

[0449] ATGGTGATTT TCGTAACAGG CGGTGCAGGA TATATTGGAT CCCATACCAT ACTTGAGTTA
[0450] CTTAATAATG GTCATGATGT CGTTTCGATA GATAATTTTG TCAATTCCTC TATAGAATCA
[0451] TTAAAAAGAG TAGAGCAAAT AACTAATAAG AAAATTATTT CTTATCAAGG TGATATCCGT
[0452] GATAAAAATC TACTTGATGA GATTTTTTCA AGACACCATA TCGATGCTGT AATTCACCTT
[0453] GCATCGTTAA AATCTGTAGG TGAGTCTAAG TTAAAGCCCT TAGAGTATTA TTCTAATAAT
[0454] GTTGGTGGA CTTTAGTATT ACTTGAATGC ATGAAGAGAT ATAACATTAA TAAAATGATA
[0455] TTTAGCTCTT CTGCTACTGT TTATGGGAGT AACAGTATCC CTCCCATAC GGAAGATAGA
[0456] CGAATTGGTG AAAC TACAAA CCCATATGGG ACATCGAAAT TTATAATAGA AATAATTTTG
[0457] AGTGATTATT GTGATAGTGA TAATAATAAA TCAGTAATTG CACTGCGTTA CTTTAATCCA
[0458] ATCGGAGCAC ATAAGTCCGG GATGATTGGT GAAAATCCTA ACGGGATCCC TAATAATCTG
[0459] GTTCCTTATA TATCTAAAGT TGCACAAAAT CAACTTCCTG TATTAAATAT TTATGGCAAC
[0460] GATTATCCAA CTAAAGATGG TACAGGAGTA AGAGACTATA TACATGTCTG TGATTTGGCT
[0461] AAAGGGCATG TTAAAGCATT AGAATATATG TTTTAAATG ATGTCAATTA TGAAGCTTTT
[0462] AATTTAGGTA CTGGTCAAGG TTATTCTGTT TTAGAGATTG TAAAAATGTT TGAGATAGTC
[0463] ACTAAAAAGA GTATACCTGT TGCTATTTGT AATAGACGTG AGGGGGATGT TCGGAGTCA
[0464] TGGGCGTCTG CTGATTTGGC ACATAAAAAG CTTTCCTGGA AAGCGAAAAA AAATTTGAAA
[0465] GAAATGATCG AAGATGTATG GCGTTGGCAA ACAACAATC CAAATGGATA TAAAAAATAA

[0466] SEQ ID NO:16-大肠杆菌086Gne的氨基酸序列

[0467] 基因座 AAV85952

[0468] 定义Gne[大肠杆菌086]。

[0469] 检索号AAV85952

[0470] 长度:339aa(gne2)线性

[0471] 类型:PRT

[0472] 生物体:大肠杆菌086

[0473] 序列:

[0474] mvifvtggag yigshtilel lnghdvvs dnfvnssies lkrveqitnk kiisyqgdir
[0475] dknlldeifs rhhidavihf asksvgesk lkpleysnn vggtlvllec mkryninkmi
[0476] fsssatvygs nsipphtedr rigettnpyg tskfiieil sdycdsdnk svialryfnp
[0477] igahksgmig enpngipnnl vpyiskvaqn qlpvl niygn dyptkdgtgv rdyihvcdla
[0478] kghvkaleym flndvnyeaf nlgtgqgysv leivkmfeiv tkksipvaic nrregdvaes
[0479] wasadlahkk lswkaeknlk emiedvwrwq tnnpngykk

[0480] SEQ ID NO:17-编码Z3206的末端的合成寡核苷酸Z3206-Fw(引物)的核苷酸

[0481] 序列;限制位点加下画线

[0482] 长度:30

[0483] 类型:DNA

[0484] 序列:AAACCCGGGATGAACGATAACGTTTTGCTC

[0485] SEQ ID NO:18-编码Z3206的末端的合成寡核苷酸Z3206-RvHA(引物)并具有血凝素标签(HA标签)的核苷酸序列;限制位点加下画线

- [0486] 长度:60
[0487] 类型:DNA
[0488] 生物体:
[0489] 序列:AAATCTAGATTAAGCGTAATCTGGAAC
[0490] ATCGTATGGGTACTCAGAAACAAACGTTATGTC
[0491] SEQ ID NO:19-合成寡核苷酸gne-Fw(引物)的核苷酸序列,其中限制位点加下画线
[0492] 长度:29
[0493] 类型:DNA
[0494] 生物体:
[0495] 序列:AAACCATGGATGAAAATTCTTATTAGCGG
[0496] SEQ ID NO:20-合成寡核苷酸gne-RV(引物)的核苷酸序列,其中限制位点加下画线
[0497] 长度:57
[0498] 类型:DNA
[0499] 生物体:
[0500] 序列:AAATCTAGATTAAGCGTAATCTGGAAC
[0501] ATCGTATGGGTAGCACTGTTTTTCCCAATC
[0502] SEQ ID NO:21-含有NheI限制酶的限制位点的寡核苷酸的核苷酸序列
[0503] 长度:11
[0504] 类型:DNA
[0505] 生物体:
[0506] 序列:AAAAAGCTAGC
[0507] SEQ ID NO:22-含有AscI限制酶的限制位点的寡核苷酸的核苷酸序列
[0508] 长度:8
[0509] 类型:DNA
[0510] 生物体:
[0511] 序列:CCGCGCGG
[0512] SEQ ID NO:23-编码Z3206的质粒pMLBAD:Z3206(质粒中的大肠杆菌0157插入片段)并具有C-端血凝素标签的核苷酸序列
[0513] 定义 产物连接到Z3206-pMLBAD*
[0514] 特征 位置/合格者
[0515] CDS 2105..3098
[0516] /标记物=Z3206
[0517] CDS 3098..3127
[0518] /标记物=HA
[0519] 长度:7794bp
[0520] 类型:DNA 环状UNA
[0521] 序列:

[0522] 1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC CCAGTTGACA
[0523] 61 TAAGCCTGTT CGGTTTCGTAA ACTGTAATGC AAGTAGCGTA TCGCTCACG CAACTGGTCC
[0524] 121 AGAACCTTGA CCGAACGCAG CGGTGGTAAC GCGCAGTGG CGGTTTTTCAT GGCTTGTTAT
[0525] 181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT AGCTAAGCGC GTTACGCCGT
[0526] 241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG CAGCAGGGTA GTCGCCCTAA
[0527] 301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG TTGTTGTGGG GTAGGCAGTC
[0528] 361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTCGACA TGGGTCAAAG
[0529] 421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTGCG CTTCCTCTGA GTGCCACCTT
[0530] 481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT TGGCAGGGTC AAGTCGTCGG
[0531] 541 TTGGTATTGC AAAAACTCA CTCCTGAAGG CTATGCGGTC GAGTCCGAAT CCCACCCAGG
[0532] 601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG GCCTAAGCGA TATCTTAGGA
[0533] 661 TCTCCCATCG GTGATGTCCG CGATATAGGC GCCAGCAACC GCACCTGTGG CGCCGGTGAT
[0534] 721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT TTGACAGCTT ATCATCGATG
[0535] 781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA AACCTATGC TACTCCGTCA
[0536] 841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA TCATTCACTT
[0537] 901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC GGTGCATTTT TTAAATACCC
[0538] 961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC GACGGTGGCG ATAGGCATCC
[0539] 1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG GTCCTCGCGC CAGCTTAAGA
[0540] 1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG CGACGGCGAC AAGCAAACAT
[0541] 1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG GTGATCGCTG ATGTACTGAC
[0542] 1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA CTCGTTAATC GCTTCCATGC
[0543] 1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG CTCCGAATAG CGCCCTTCCC
[0544] 1321 CTTGCCCCGGC GTTAATGATT TGCCCCAACA GGTGCTGAA ATGCGGCTGG TGCGCTTCAT
[0545] 1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA GTTAAGCCAT TCATGCCAGT
[0546] 1441 AGGCGCGCGG ACGAAAAGTAA ACCCACTGGT GATACCATTG GCGAGCCTCC GGATGACGAC
[0547] 1501 CGTAGTGATG AATCTCTCCT GGCGGGAACA GCAAAATATC ACCCGGTCGG CAAACAAATT
[0548] 1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT GAGATTGAGA ATATAACCTT
[0549] 1621 TCATTCCCAG CGGTCGGTCG ATAAAAAAT CGAGATAACC GTTGGCCTCA ATCGGCGTTA
[0550] 1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG CAGGGGATCA TTTTGCCTT
[0551] 1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAGAAGAAAC CAATTGTCCA TATTGCATCA
[0552] 1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT AACCAAAACG GTAACCCCGC
[0553] 1861 TTATTAAAAG CATTCTGTAA CAAAGCGGGA CCAAAGCCAT GACAAAAACG CGTAACAAAA
[0554] 1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT GCACGGCGTC ACACTTTGCT
[0555] 1981 ATGCCATAGC ATTTTATATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT TTATCGCAAC
[0556] 2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG GAGGAATTCA CCATGGTACC
[0557] 2101 CGGGATGAAC GATAACGTTT TGCTCATAGG AGCTTCCGGA TTCGTAGGAA CCCGACTACT
[0558] 2161 TGAAACGGCA ATTGCTGACT TTAATATCAA GAACCTGGAC AAACAGCAGA GCCACTTTTA
[0559] 2221 TCCAGAAATC ACACAGATTG GCGATGTTTCG CGATCAACAG GCACTCGACC AGGCGTTAGT
[0560] 2281 CGGTTTTGAC ACTGTTGTAC TACTGGCAGC GGAACACCGC GATGACGTCA GCCCTACTTC

[0561] 2341 TCTCTATTAT GATGTCAACG TTCAGGGTAC CCGCAATGTG CTGGCGGCCA TGGAAAAAAA
[0562] 2401 TGGCGTAAA AATATCATCT TTACCAGTTC CGTTGCTGTT TATGGTTTGA ACAAACACAA
[0563] 2461 CCCTGACGAA AACCATCCAC ACGACCCTTT CAACCACTAC GGCAAAAAGTA AGTGGCAGGC
[0564] 2521 AGAGGAAGTG CTGCGTGAAT GGTATAACAA AGCACCAACA GAACGTTTCAT TAACCATCAT
[0565] 2581 CCGTCCTACC GTTATCTTCG GTGAACGCAA CCGCGGTAAC GTCTATAACT TGCTGAAACA
[0566] 2641 GATCGCTGGC GGCAAGTTTA TGATGGTGGG CGCAGGGACT AACTATAAGT CCATGGCTTA
[0567] 2701 TGTGGAAAC ATTGTTGAGT TTATCAAGTA CAACTGAAG AATGTTGCCG CAGGTTATGA
[0568] 2761 GGTTTATAAC TACGTTGATA AGCCAGACCT GAACATGAACCAGTTGGTTGC TGAAGTTGA
[0569] 2821 ACAAAGCCTG AACAAAAAGA TCCCTTCTAT GCACTTGCCT TACCACTAG GAATGCTGGG
[0570] 2881 TGGATATTGC TTTGATATCC TGAGCAAAAT TACGGGCAAA AAATACGCTG TCAGCTCAGT
[0571] 2941 GCGCGTGAAG AAATCTGCG CAACAACACA GTTTGACGCA ACGAAAGTGC ATTCTTCAGG
[0572] 3001 TTTTGTGGCA CCGTATACGC TGTCGCAAGG TCTGGATCGA AACTGTCAGT ATGAATTCGT
[0573] 3061 TCATGCCAAA AAAGACGACA TAACGTTTGT TTCTGAGTAC CCATACGATG TTCCAGATTA
[0574] 3121 CGCTTAATCT AGAGTCGACC TGCAGGCATG CAAGCTTGGC TGTTTTGGCG GATGAGAGAA
[0575] 3181 GATTTTCAGC CTGATACAGA TTAAATCAGA ACGCAGAAGC GGTCTGATAA AACAGAATTT
[0576] 3241 GCCTGGCGGC AGTAGCGCGG TGGTCCCACC TGACCCCATG CCGAACTCAG AAGTGAAACG
[0577] 3301 CCGTAGCGCC GATGGTAGTG TGGGGTCTCC CCATGCGAGA GTAGGGAAGT GCCAGGCATC
[0578] 3361 AAATAAAACG AAAGGCTCAG TCGAAAGACT GGGCCTTTTCG TTTTATCTGT TGTTTGTGCG
[0579] 3421 TGAACGCTCT CCTGAGTAGG ACAAATCCGC CGGGAGCGGA TTTGAACGTT GCGAAGCAAC
[0580] 3481 GGCCCGGAGG GTGGCGGGCA GGACGCCCGC CATAAACTGC CAGGCATCAA ATTAAGCAGA
[0581] 3541 AGGCCATCCT GACGGATGGC CTTTTTTCGT TTCTACAAAC TCTTCCACTC ACTACAGCAG
[0582] 3601 AGCCATTTAA ACAACATCCC CTCCCCCTTT CCACCGCGTC AGACGCCCGT AGCAGCCCGC
[0583] 3661 TACGGGCTTT TTCATGCCCT GCCCTAGCGT CCAAGCCTCA CGGCCGCGCT CGGCCTCTCT
[0584] 3721 GGCGGCCTTC TGGCGCTGAG GTCTGCCTCG TGAAGAAGGT GTTGCTGACT CATACCAGGC
[0585] 3781 CTGAATCGCC CCATCATCCA GCCAGAAAGT GAGGGAGCCA CGGTTGATGA GAGCTTTGTT
[0586] 3841 GTAGGTGGAC CAGTTGGTGA TTTTGAAGTT TTGCTTTGCC ACGGAACGGT CTGCGTTGTC
[0587] 3901 GGGAAGATGC GTGATCTGAT CCTTCAACTC AGCAAAAGTT CGATTTATTC AACAAAGCCG
[0588] 3961 CCGTCCCGTC AAGTCAGCGT AATGCTCTGC CAGTGTTACA ACCAATTAAC CAATTCTGAT
[0589] 4021 TAGAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT TCATATCAGG ATTATCAATA
[0590] 4081 CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA ACTCACCAGG GCAGTTCCAT
[0591] 4141 AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC GTCCAACATC AATACAACCT
[0592] 4201 ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGCGAGA AATCACCATG AGTGACGACT
[0593] 4261 GAATCCGGTG AGAATGGCAA AAGCTAAAAA GGCCGTAATA TCCAGCTGAA CCGTCTGGTT
[0594] 4321 ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAAATGT TCTTTACGAT GCCATTGGGA
[0595] 4381 TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT TAGCTCCTGA
[0596] 4441 AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTCAATTAT GGTGAAAGTT
[0597] 4501 GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC AGGGCTTCCC
[0598] 4561 GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG TGATCTTCG TCACAGGTAT
[0599] 4621 TTATTCGAAG ACGAAAGGGC CTCGTGATAC GCCTATTTTT ATAGGTTAAT GTCATGATAA

[0600] 4681 TAATGGTTTC TTAGACGTCA GGTGGCACTT TTCGGGAAA TGTGCGCGCC CGCGTTCCTG
[0601] 4741 CTGGCGCTGG GCCTGTTTCT GGCCTGGAC TTCCCGCTGT TCCGTCAGCA GCTTTTCGCC
[0602] 4801 CACGGCCTTG ATGATCGCGG CGGCCTTGGC CTGCATATCC CGATTCAACG GCCCCAGGGC
[0603] 4861 GTCCAGAACG GGCTTCAGGC GCTCCGAAG GTCTCGGGCC GTCTCTTGGG CTTGATCGGC
[0604] 4921 CTTCTTGCGC ATCTCACGCG CTCCTGCGGC GGCCTGTAGG GCAGGCTCAT ACCCCTGCCG
[0605] 4981 AACCGCTTTT GTCAGCCGGT CGGCCACGGC TTCCGGCGTC TCAACGCGCT TTGAGATTCC
[0606] 5041 CAGCTTTTCG GCCAATCCCT GCGGTGCATA GGCGCGTGGC TCGACCGCTT GCGGGCTGAT
[0607] 5101 GGTGACGTGG CCCACTGGTG GCCGCTCCAG GGCCTCGTAG AACGCTGAA TGCGCGTGTG
[0608] 5161 ACGTGCCTTG CTGCCCTCGA TGCCCCGTTG CAGCCCTAGA TCGGCCACAG CGGCCGCAAA
[0609] 5221 CGTGGTCTGG TCGCGGGTCA TCTGCGCTTT GTTGCCGATG AACTCCTTGG CCGACAGCCT
[0610] 5281 GCCGTCCTGC GTCAGCGGCA CCACGAACGC GGTCATGTGC GGGCTGGTTT CGTCACGGTG
[0611] 5341 GATGCTGGCC GTCACGATGC GATCCGCCCC GTACTTGTCC GCCAGCCACT TGTGCGCCTT
[0612] 5401 CTCGAAGAAC GCCGCTGCT GTTCTTGGCT GGCCGACTTC CACCATTCCG GGCTGGCCGT
[0613] 5461 CATGACGTAC TCGACCGCCA ACACAGCGTC CTTGCGCCGC TTCTCTGGCA GCAACTCGCG
[0614] 5521 CAGTCGGCCC ATCGCTTCAT CGGTGCTGCT GGCCGCCCAG TGCTCGTTCT CTGGCGTCTT
[0615] 5581 GCTGGCGTCA GCGTTGGGCG TCTCGCGCTC GCGGTAGGCG TGCTTGAGAC TGGCCGCCAC
[0616] 5641 GTTGCCCAT TTCGCCAGCT TCTTGATCG CATGATCGCG TATGCCGCCA TGCTGCCCC
[0617] 5701 TCCCTTTTGG TGTCCAACCG GCTCGACGGG GGCAGCGCAA GGCGGTGCCT CCGGCGGGCC
[0618] 5761 ACTCAATGCT TGAGTATACT CACTAGACTT TGCTTCGCAA AGTCGTGACC GCCTACGGCG
[0619] 5821 GCTGCGGCGC CCTACGGGCT TGCTCTCCGG GCTTCGCCCT GCGCGGTGCG TGCGTCCCT
[0620] 5881 TGCCAGCCCG TGGATATGTG GACGATGGCC GCGAGCGGCC ACCGGCTGGC TCGCTTCGCT
[0621] 5941 CGGCCCCGTG ACAACCCCTGC TGGACAAGCT GATGGACAGG CTGCGCCTGC CCACGAGCTT
[0622] 6001 GACCACAGGG ATTGCCACC GGCTACCCAG CCTTCGACCA CATACCACC GGCTCCAAC
[0623] 6061 GCGCGGCCTG CGGCCTTGCC CCATCAATTT TTTAATTTT CTCTGGGGAA AAGCCTCCGG
[0624] 6121 CCTGCGGCCT GCGCGCTTCG CTTGCCGTTT GGACACCAAG TGGAAGGCGG GTCAAGGCTC
[0625] 6181 GCGCAGCGAC CGCGCAGCGG CTTGGCCTTG ACGCGCCTGG AACGACCCAA GCCTATGCGA
[0626] 6241 GTGGGGGAG TCGAAGGCGA AGCCCCCGG CCTGCCCCC GAGCCTCACG GCGGCGAGTG
[0627] 6301 CGGGGGTTCC AAGGGGGCAG CGCCACCTTG GGCAAGGCCG AAGGCCGCGC AGTCGATCAA
[0628] 6361 CAAGCCCCG AGGGGCCACT TTTTGCCGA GGGGGAGCCG CGCCGAAGGC GTGGGGGAAC
[0629] 6421 CCCGAGGGG TGCCCTTCTT TGGGCACCAA AGAACTAGAT ATAGGGCGAA ATGCGAAAGA
[0630] 6481 CTAAAAATC AACAACTTAA AAAAGGGGGG TACGCAACAG CTCATTGCGG CACCCCCCGC
[0631] 6541 AATAGCTCAT TGCGTAGGTT AAAGAAAATC TGTAATTGAC TGCCACTTTT ACGCAACGCA
[0632] 6601 TAATTGTTGT CGCGCTGCCG AAAAGTTGCA GCTGATTGCG CATGGTGCCG CAACCGTGCG
[0633] 6661 GCACCCTACC GCATGGAGAT AAGCATGGCC ACGCAGTCCA GAGAAATCGG CATTCAAGCC
[0634] 6721 AAGAACAAGC CCGGTCACTG GGTGCAAACG GAACGCAAAG CGCATGAGGC GTGGGCCGGG
[0635] 6781 CTTATTGCGA GGAAACCCAC GCGGGCAATG CTGCTGCATC ACCTCGTGGC GCAGATGGGC
[0636] 6841 CACCAGAACG CCGTGGTGGT CAGCCAGAAG AACTTTCCA AGCTCATCGG ACGTTCTTTG
[0637] 6901 CGGACGGTCC AATACGAGT CAAGGACTTG GTGGCCGAGC GCTGGATCTC CGTCGTGAAG
[0638] 6961 CTCAACGGCC CCGGCACCGT GTCGGCTAC GTGGTCAATG ACCGCGTGGC GTGGGGCCAG

[0639] 7021 CCCCCGACC AGTTGCGCCT GTCGGTGTTC AGTGCCGCCG TGGTGGTTGA TCACGACGAC

[0640] 7081 CAGGACGAAT CGCTGTTGGG GCATGGCGAC CTGCGCCGCA TCCCGACCCT GTATCCGGGC

[0641] 7141 GAGCAGCAAC TACCGACCGG CCCC GGCGAG GAGCCGCCA GCCAGCCCGG CATTCCGGGC

[0642] 7201 ATGGAACCAG ACCTGCCAGC CTTGACCGAA ACGGAGGAAT GGGAACGGCG CGGGCAGCAG

[0643] 7261 CGCCTGCCGA TGCCCGATGA GCCGTGTTTT CTGGACGATG GCGAGCCGTT GGAGCCGCCG

[0644] 7321 ACACGGGTCA CGCTGCCGCG CCGGTAGCAC TTGGGTGCG CAGCAACCCG TAAGTGCGCT

[0645] 7381 GTTCCAGACT ATCGGCTGTA GCCGCCTCGC CGCCCTATAC CTTGTCTGCC TCCCCGCGTT

[0646] 7441 GCGTCGCGGT GCATGGAGCC GGGCCACCTC GACCTGAATG GAAGCCGGCG GCACCTCGCT

[0647] 7501 AACGGATTCA CCGTTTTTAT CAGGCTCTGG GAGGCAGAAT AAATGATCAT ATCGTCAATT

[0648] 7561 ATTACCTCCA CGGGGAGAGC CTGAGCAAAC TGGCCTCAGG CATTTGAGAA GCACACGGTC

[0649] 7621 AACTGCTTC CGGTAGTCAA TAAACCGGTA AACCAGCAAT AGACATAAGC GGCTATTTAA

[0650] 7681 CGACCCTGCC CTGAACCGAC GACCGGGTCG AATTTGCTTT CGAATTTCTG CCATTCATCC

[0651] 7741 GCTTATTATC ACTTATTCAG GCGTAGCACC AGGCGTTTAA GTCGACCAAT AACC

[0652] SEQ ID NO:24-编码Gne的pMLBAD:gne(质粒中的大肠杆菌0157插入片段)并具有C-端血凝素标签的核苷酸序列

[0653] 基因座 gne-pMLBAD

[0654] 定义dig gale连接到pmlbad did(NcoI-XbaI)

[0655] 特征 位置/合格者

[0656] CDS 2097..3080

[0657] /标记物=gale

[0658] CDS 3081..3107

[0659] /标记物=HA

[0660] 区域 3108..3110

[0661] /标记物=stop

[0662] 长度:7776bp

[0663] 类型:DNA 环状UNA

[0664] 序列:

[0665] 1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC CCAGTTGACA

[0666] 61 TAAGCCTGTT CGGTTTCGTAA ACTGTAATGC AAGTAGCGTA TCGCTCACG CAACTGGTCC

[0667] 121 AGAACCTTGA CCGAACGCAG CCGTGGAAC GCGCAGTGG CGGTTTTCAT GGCTTGTTAT

[0668] 181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT AGCTAAGCGC GTTACGCCGT

[0669] 241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG CAGCAGGGTA GTCGCCCTAA

[0670] 301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG TTGTTGTGGG GTAGGCAGTC

[0671] 361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTGACA TGGGTCAAAG

[0672] 421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTTCG CTCCCCTGA GTGCCACCTT

[0673] 481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT TGGCAGGGTC AAGTCGTCGG

[0674] 541 TTGGTATTGC AAAAACTCA CTCCTGAAGG CTATGCGGTC GAGTCCGAAT CCCACCCAGG

[0675] 601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG GCCTAAGCGA TATCTTAGGA

[0676] 661 TCTCCCATCG GTGATGTGCG CGATATAGGC GCCAGCAACC GCACCTGTGG CGCCGGTGAT

[0677] 721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT TTGACAGCTT ATCATCGATG
[0678] 781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA AACCTATGC TACTCCGTCA
[0679] 841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA TCATTCACTT
[0680] 901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC GGTGCATTTT TTAAATACCC
[0681] 961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC GACGGTGGCG ATAGGCATCC
[0682] 1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG GTCTCGCGC CAGCTTAAGA
[0683] 1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG CGACGGCGAC AAGCAAACAT
[0684] 1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG GTGATCGCTG ATGTACTGAC
[0685] 1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA CTCGTTAATC GCTTCCATGC
[0686] 1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG CTCCGAATAG CGCCCTTCCC
[0687] 1321 CTTGCCCCGGC GTTAATGATT TGCCCCAAACA GGTGCTGAA ATGCGGCTGG TGCGCTTCAT
[0688] 1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA GTTAAGCCAT TCATGCCAGT
[0689] 1441 AGGCGCGCGG ACGAAAGTAA ACCCACTGGT GATACCATTG GCGAGCCTCC GGATGACGAC
[0690] 1501 CGTAGTGATG AATCTCTCCT GGCGGGAACA GCAAAATATC ACCCGGTGCG CAAACAAATT
[0691] 1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT GAGATTGAGA ATATAACCTT
[0692] 1621 TCATTCCCAG CGGTGCGTCG ATAAAAAAT CGAGATAACC GTTGGCCTCA ATCGGCGTTA
[0693] 1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG CAGGGGATCA TTTTGCGCTT
[0694] 1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAGAAGAAAC CAATTGTCCA TATTGCATCA
[0695] 1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT AACCAAACCG GTAACCCCGC
[0696] 1861 TTATTAAG CATTCTGTAA CAAAGCGGA CCAAAGCCAT GACAAAAACG CGTAACAAAA
[0697] 1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT GCACGGCGTC ACACCTTGCT
[0698] 1981 ATGCCATAGC ATTTTATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT TTATCGCAAC
[0699] 2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG GAGGAATTCA CCATGGATGA
[0700] 2101 AAATTCTTAT TAGCGGTGGT GCAGGTTATA TAGGTTCTCA TACTTTAAGA CAATTTTAA
[0701] 2161 AAACAGATCA TGAAATTTGT GTTTTAGATA ATCTTTCTAA GGGTTCTAAA ATCGCAATAG
[0702] 2221 AAGATTTGCA AAAAATAAGA ACTTTTAAAT TTTTGAACA AGATTTAAGT GATTTTCAAG
[0703] 2281 GCGTAAAAGC ATTGTTTGAG AGAGAAAAAT TTGACGCTAT TGTGCATTTT GCAGCGAGCA
[0704] 2341 TTGAAGTTTT TGAAAGTATG CAAAACCCTT TAAAGTATTA TATGAATAAC ACTGTTAATA
[0705] 2401 CGACAAATCT CATCGAAACT TGTTTGCAA CTGGAGTGAA TAAATTTATA TTTTCTTCAA
[0706] 2461 CGGCAGCCAC TTATGGCGAA CCACAACTC CCGTTGTGAG CGAAACAAGT CCTTTAGCAC
[0707] 2521 CTATTAATCC TTATGGGCGT AGTAAGCTTA TGAGCGAAGA GGTTTTGCGT GATGCAAGTA
[0708] 2581 TGGCAAATCC TGAATTTAAG CATTGTATTT TAAGATATTT TAATGTTGCA GGTGCTTGCA
[0709] 2641 TGGATTATAC TTTAGGACAA CGCTATCCAA AAGCGACTTT GCTTATAAAA GTTGCAGCTG
[0710] 2701 AATGTGCCGC AGAAAAACGT AATAAACTTT TCATATTTGG CGATGATTAT GATACAAAAG
[0711] 2761 ATGGCACTTG CATAAGAGAT TTTATCCATG TGGATGATAT TTCAAGTGCG CATTTATCGG
[0712] 2821 CTTTGGATTA TTTAAAAGAG AATGAAAGCA ATGTTTTTAA TGTAGGTTAT GGACATGGTT
[0713] 2881 TTAGCGTAAA AGAAGTGATT GAAGCGATGA AAAAAGTTAG CGGAGTGGAT TTTAAAGTAG
[0714] 2941 AACTTGCCCC ACGCCGTGCG GGTGATCCTA GTGTATTGAT TTCTGATGCA AGTAAATCA
[0715] 3001 GAAATCTTAC TTCTTGGCAG CCTAAATATG ATGATTTAGG GCTTATTTGT AAATCTGCTT

[0716] 3061 TTGATTGGGA AAAACAGTGC TACCCATACG ATGTTCCAGA TTACGCTTAA TCTAGAGTCG
 [0717] 3121 ACCTGCAGGC ATGCAAGCTT GGCTGTTTTG GCGGATGAGA GAAGATTTTC AGCCTGATAC
 [0718] 3181 AGATTAAATC AGAACGCAGA AGCGGTCTGA TAAAACAGAA TTTGCCTGGC GGCAGTAGCG
 [0719] 3241 CGGTGGTCCC ACCTGACCCC ATGCCGAAC TACAAGTGAA ACGCCGTAGC GCCGATGGTA
 [0720] 3301 GTGTGGGGTC TCCCCATGCG AGAGTAGGGA ACTGCCAGGC ATCAAATAAA ACGAAAGGCT
 [0721] 3361 CAGTCGAAAG ACTGGGCCTT TCGTTTTATC TGTGTGTTGT CGGTGAACGC TCTCCTGAGT
 [0722] 3421 AGGACAAATC CGCCGGGAGC GGATTTGAAC GTTGCGAAGC AACGGCCCCG AGGGTGGCGG
 [0723] 3481 GCAGGACGCC CGCCATAAAC TGCCAGGCAT CAAATTAAGC AGAAGGCCAT CCTGACGGAT
 [0724] 3541 GGCCTTTTTG CGTTTCTACA AACTCTTCCA CTCACTACAG CAGAGCCATT TAAACAACAT
 [0725] 3601 CCCCTCCCCC TTTCCACCGC GTCAGACGCC CGTAGCAGCC CGCTACGGGC TTTTTCATGC
 [0726] 3661 CCTGCCCTAG CGTCCAAGCC TCACGGCCGC GCTCGGCCCTC TCTGGCGGCC TTCTGGCGCT
 [0727] 3721 GAGGTCTGCC TCGTGAAGAA GGTGTTGCTG ACTCATACCA GGCCTGAATC GCCCCATCAT
 [0728] 3781 CCAGCCAGAA AGTGAGGGAG CCACGGTTGA TGAGAGCTTT GTTGTAGGTG GACCAGTTGG
 [0729] 3841 TGATTTTGAA CTTTGTCTTT GCCACGGAAC GGTCTGCGTT GTCGGGAAGA TGCGTGATCT
 [0730] 3901 GATCCTTCAA CTCAGCAAAA GTTCGATTTA TTCAACAAAG CCGCCGTCCC GTCAAGTCAG
 [0731] 3961 CGTAATGCTC TGCCAGTGTT ACAACCAATT AACCAATTCT GATTAGAAAA ACTCATCGAG
 [0732] 4021 CATCAAATGA AACTGCAATT TATTCATATC AGGATTATCA ATACCATATT TTTGAAAAAG
 [0733] 4081 CCGTTTCTGT AATGAAGGAG AAAACTCACC GAGGCAGTTC CATAGGATGG CAAGATCCTG
 [0734] 4141 GTATCGGTCT GCGATTCCGA CTCGTCCAAC ATCAATACAA CCTATTAATT TCCCCTCGTC
 [0735] 4201 AAAAATAAGG TTATCAAGCG AGAAATCACC ATGAGTGACG ACTGAATCCG GTGAGAATGG
 [0736] 4261 CAAAAGCTAA AAAGGCCGTA ATATCCAGCT GAACGGTCTG GTTATAGGTA CATTGAGCAA
 [0737] 4321 CTGACTGAAA TGCCTCAAAA TGTTCTTTAC GATGCCATTG GGATATATCA ACGGTGGTAT
 [0738] 4381 ATCCAGTGAT TTTTTTCTCC ATTTTAGCTT CCTTAGCTCC TGAANAATCTC GATAACTCAA
 [0739] 4441 AAAATACGCC CGGTAGTGAT CTTATTTTAT TATGGTGAAA GTTGGAACCT CTTACGTGCC
 [0740] 4501 GATCAACGTC TCATTTTCGC CAAAAGTTGG CCCAGGGCTT CCCGGTATCA ACAGGGACAC
 [0741] 4561 CAGGATTTAT TTATTCTGCG AAGTGATCTT CCGTCACAGG TATTTATTCTG AAGACGAAAG
 [0742] 4621 GGCCTCGTGA TACGCCTATT TTTATAGGTT AATGTCATGA TAATAATGGT TTCTTAGACG
 [0743] 4681 TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GCCCGCGTTC CTGCTGGCGC TGGGCCTGTT
 [0744] 4741 TCTGGCGCTG GACTTCCCGC TGTTCCGTCA GCAGCTTTTC GCCACGGCC TTGATGATCG
 [0745] 4801 CGGCGGCCCTT GGCCTGCATA TCCCGATTCA ACGGCCCCAG GCGTCCAGA ACGGGCTTCA
 [0746] 4861 GGCCTCCCG AAGGTCTCGG GCCGTCTCTT GGGCTTGATC GGCCTTCTTG CGCATCTCAC
 [0747] 4921 GCGCTCCTGC GCGGCCTGT AGGGCAGGCT CATACCCCTG CCGAACCCTG TTTGTGAGCC
 [0748] 4981 GGTGCGCCAC GGCTCCGGC GTCTCAACGC GCTTTGAGAT TCCCAGCTTT TCGGCCAATC
 [0749] 5041 CCTGCGGTGC ATAGGCGCGT GGCTCGACCG CTGCGGGCT GATGGTGACG TGGCCACTG
 [0750] 5101 GTGGCCGCTC CAGGGCCTCG TAGAACGCCT GAATGCGCGT GTGACGTGCC TTGCTGCCCT
 [0751] 5161 CGATGCCCCG TTGCAGCCCT AGATCGGCCA CAGCGGCCGC AAACGTGGTC TGGTCGCGGG
 [0752] 5221 TCATCTGCGC TTTGTTGCCG ATGAACTCCT TGGCCGACAG CCTGCCGTCC TCGCTCAGCG
 [0753] 5281 GCACCACGAA CGCGGTCATG TGCGGGCTGG TTTCGTCACG GTGGATGCTG GCCGTACGA
 [0754] 5341 TGCGATCCGC CCCGTACTTG TCCGCCAGCC ACTTGTGCGC CTCTCGAAG AACGCCGCT

[0755] 5401 GCTGTTCTTG GCTGGCCGAC TTCCACCATT CCGGGCTGGC CGTCATGACG TACTCGACCG
[0756] 5461 CCAACACAGC GTCCTTGCGC CGCTTCTCTG GCAGCAACTC GCGCAGTCGG CCCATCGCTT
[0757] 5521 CATCGGTGCT GCTGGCCGCC CAGTGCTCGT TCTCTGGCGT CCTGCTGGCG TCAGCGTTGG
[0758] 5581 GCGTCTCGCG CTCGCGGTAG GCGTGCTTGA GACTGGCCGC CACGTTGCCC ATTTTCGCCA
[0759] 5641 GCTTCTTGCA TCGCATGATC GCGTATGCCG CCATGCCTGC CCCTCCCTTT TGGTGTCCAA
[0760] 5701 CCGGCTCGAC GGGGGCAGCG CAAGGCGGTG CCTCCGGCGG GCCACTCAAT GCTTGAGTAT
[0761] 5761 ACTACTAGA CTTTGCTTCG CAAAGTCGTG ACCGCCTACG GCGGCTGCGG CGCCCTACGG
[0762] 5821 GCTTGCTCTC CGGGCTTCGC CCTGCGCGGT CGCTGCGCTC CCTTGCCAGC CCGTGGATAT
[0763] 5881 GTGGACGATG GCCGCGAGCG GCCACCGGT GGCTCGCTTC GCTCGGCCCG TGGACAACCC
[0764] 5941 TGCTGGACAA GCTGATGGAC AGGCTGCGCC TGCCCACGAG CTTGACCACA GGGATTGCCC
[0765] 6001 ACCGGCTACC CAGCCTTCGA CCACATACCC ACCGGCTCCA ACTGCGCGGC CTGCGGCCTT
[0766] 6061 GCCCCATCAA TTTTTTTAAT TTTCTCTGGG GAAAAGCCTC CGGCCTGCGG CCTGCGCGCT
[0767] 6121 TCGCTTGCCG GTTGACACC AAGTGGAAG CGGGTCAAG CTCGCGCAGC GACCGCGCAG
[0768] 6181 CGGCTTGGCC TTGACGCGCC TGGAACGACC CAAGCCTATG CGAGTGGGGG CAGTCGAAGG
[0769] 6241 CGAAGCCCGC CCGCCTGCCC CCCGAGCCTC ACGGCGGCGA GTGCGGGGGT TCCAAGGGGG
[0770] 6301 CAGCGCCACC TTGGGCAAGG CCGAAGGCCG CGCAGTCGAT CAACAAGCCC CGGAGGGGCC
[0771] 6361 ACTTTTTGCC GGAGGGGGAG CCGCGCCGAA GCGTGGGGG AACCCCGCAG GGGTGCCCTT
[0772] 6421 CTTTGGGCAC CAAAGAACTA GATATAGGGC GAAATGCGAA AGACTTAAAA ATCAACAAC
[0773] 6481 TAAAAAAGG GGTACGCAA CAGCTCATTG CGGCACCCCC CGCAATAGCT CATTGCGTAG
[0774] 6541 GTTAAAGAAA ATCTGTAATT GACTGCCACT TTTACGCAAC GCATAATTGT TGTCGCGCTG
[0775] 6601 CCGAAAAGTT GCAGCTGATT GCGCATGGTG CCGCAACCGT GCGGCACCCCT ACCGCATGGA
[0776] 6661 GATAAGCATG GCCACGCAGT CCAGAGAAAT CGGCATTCAA GCCAAGAACA AGCCCGGTCA
[0777] 6721 CTGGGTGCAA ACGGAACGCA AAGCGCATGA GCGTGGGGC GGGCTTATTG CGAGGAAACC
[0778] 6781 CACGGCGGCA ATGCTGCTGC ATCACCTCGT GGCGCAGATG GGCCACCAGA ACGCCGTGGT
[0779] 6841 GGTACGCCAG AAGACACTTT CCAAGCTCAT CGGACGTTCT TTGCGGACGG TCCAATACGC
[0780] 6901 AGTCAAGGAC TTGGTGCCG AGCGCTGGAT CTCCGTCTGT AAGCTCAACG GCCCGGCAC
[0781] 6961 CGTGTCGGCC TACGTGGTCA ATGACCGGT GCGTGGGGC CAGCCCCGCG ACCAGTTGCG
[0782] 7021 CCTGTGCGTG TTCAGTGCCG CCGTGGTGGT TGATCACGAC GACCAGGACG AATCGCTGTT
[0783] 7081 GGGGCATGGC GACCTGCGCC GCATCCCGAC CCTGTATCCG GGCGAGCAGC AACTACCGAC
[0784] 7141 CGGCCCCGGC GAGGAGCCGC CCAGCCAGCC CGGCATTCCG GGCATGGAAC CAGACCTGCC
[0785] 7201 AGCCTTGACC GAAACGGAGG AATGGGAACG GCGCGGGCAG CAGCGCCTGC CGATGCCCGA
[0786] 7261 TGAGCCGTGT TTTCTGGACG ATGGCGAGCC GTTGGAGCCG CCGACACGGG TCACGCTGCC
[0787] 7321 GCGCCGGTAG CACTTGGGTT GCGCAGCAAC CCGTAAGTGC GCTGTTCCAG ACTATCGGCT
[0788] 7381 GTAGCCGCCT CGCCGCCCTA TACCTTGTCT GCCTCCCCGC GTTGCCTGCG GGTGCATGGA
[0789] 7441 GCCGGGCCAC CTCGACCTGA ATGGAAGCCG GCGGCACCTC GCTAACGGAT TCACCGTTTT
[0790] 7501 TATCAGGCTC TGGGAGGCAG AATAAATGAT CATATCGTCA ATTATTACCT CCACGGGGAG
[0791] 7561 AGCCTGAGCA AACTGGCCTC AGGCATTGTA GAAGCACACG GTCACACTGC TTCCGGTAGT
[0792] 7621 CAATAAACCG GTAAACCAGC AATAGACATA AGCGGCTATT TAACGACCCT GCCCTGAACC
[0793] 7681 GACGACCGGG TCGAATTTGC TTTCGAATTT CTGCCATTCA TCCGCTTATT ATCACTTATT

- [0794] 7741 CAGGCGTAGC ACCAGGCGTT TAAGTCGACC AATAAC
[0795] SEQ ID NO:25-具有信号序列的修饰EPA的氨基酸序列
[0796] 公开于WO 2009/104074(作为SEQ ID NO.6)
[0797] 类型:PRT
[0798] 生物体:人工的
[0799] /注释="人工序列的描述:合成多肽"

[0800] 长度:643

[0801] 序列:

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
1 5 10 15

Ala Ser Ala Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys
20 25 30

Ala Cys Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser
35 40 45

[0802] Val Asp Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr
50 55 60

Ser Met Val Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp
65 70 75 80

Asn Ala Leu Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly
85 90 95

Gly Val Glu Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala
100 105 110

Arg Gly Ser Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys
115 120 125

Pro Ser Asn Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln
130 135 140

Leu Ser His Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu
145 150 155 160

Leu Ala Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu
165 170 175

Ser Asn Glu Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser
180 185 190

[0803] Val Val Met Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu
195 200 205

Trp Ala Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val
210 215 220

Tyr Asn Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu
225 230 235 240

Gly Lys Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu
245 250 255

Asp Ile Lys Asp Asn Asn Asn Ser Thr Pro Thr Val Ile Ser His Arg
260 265 270

Leu His Phe Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln
275 280 285

Ala Cys His Leu Pro Leu Glu Ala Phe Thr Arg His Arg Gln Pro Arg
290 295 300

Gly Trp Glu Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val
305 310 315 320

Ala Leu Tyr Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val
325 330 335

Ile Arg Asn Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu
340 345 350

Ala Ile Arg Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala
355 360 365

[0804] Ala Ala Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu
370 375 380

Ala Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala
385 390 395 400

Lys Asp Gln Asn Arg Thr Lys Gly Glu Cys Ala Gly Pro Ala Asp Ser
405 410 415

Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu
420 425 430

Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp
435 440 445

Thr Val Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly
450 455 460

Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser
465 470 475 480

Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile
485 490 495

Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr
500 505 510

Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala
515 520 525

Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg
530 535 540

[0805] Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu Arg
545 550 555 560

Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro
565 570 575

Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala Glu
580 585 590

Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val
595 600 605

Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile
610 615 620

Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu
625 630 635 640

[0806] Asp Leu Lys

[0807] SEQ ID NO:26-Pg1B的氨基酸序列

[0808] 公开于W0 2009/104074(作为SEQ ID NO.2)

[0809] 长度:722

[0810] 类型:PRT

[0811] 生物体:空肠弯曲杆菌(*Campylobacter jejuni*)

[0812] 序列:

Met Leu Lys Lys Glu Tyr Leu Lys Asn Pro Tyr Leu Val Leu Phe Ala
1 5 10 15

Met Ile Ile Leu Ala Tyr Val Phe Ser Val Phe Cys Arg Phe Tyr Trp
20 25 30

Val Trp Trp Ala Ser Glu Phe Asn Glu Tyr Phe Phe Asn Asn Gln Leu
35 40 45

[0813] Met Ile Ile Ser Asn Asp Gly Tyr Ala Phe Ala Glu Gly Ala Arg Asp
50 55 60

Met Ile Ala Gly Phe His Gln Pro Asn Asp Leu Ser Tyr Tyr Gly Ser
65 70 75 80

Ser Leu Ser Ala Leu Thr Tyr Trp Leu Tyr Lys Ile Thr Pro Phe Ser
85 90 95

Phe Glu Ser Ile Ile Leu Tyr Met Ser Thr Phe Leu Ser Ser Leu Val
100 105 110

Val Ile Pro Thr Ile Leu Leu Ala Asn Glu Tyr Lys Arg Pro Leu Met
115 120 125

Gly Phe Val Ala Ala Leu Leu Ala Ser Ile Ala Asn Ser Tyr Tyr Asn
130 135 140

Arg Thr Met Ser Gly Tyr Tyr Asp Thr Asp Met Leu Val Ile Val Leu
145 150 155 160

Pro Met Phe Ile Leu Phe Phe Met Val Arg Met Ile Leu Lys Lys Asp
165 170 175

Phe Phe Ser Leu Ile Ala Leu Pro Leu Phe Ile Gly Ile Tyr Leu Trp
180 185 190

Trp Tyr Pro Ser Ser Tyr Thr Leu Asn Val Ala Leu Ile Gly Leu Phe
195 200 205

[0814]

Leu Ile Tyr Thr Leu Ile Phe His Arg Lys Glu Lys Ile Phe Tyr Ile
210 215 220

Ala Val Ile Leu Ser Ser Leu Thr Leu Ser Asn Ile Ala Trp Phe Tyr
225 230 235 240

Gln Ser Ala Ile Ile Val Ile Leu Phe Ala Leu Phe Ala Leu Glu Gln
245 250 255

Lys Arg Leu Asn Phe Met Ile Ile Gly Ile Leu Gly Ser Ala Thr Leu
260 265 270

Ile Phe Leu Ile Leu Ser Gly Gly Val Asp Pro Ile Leu Tyr Gln Leu
275 280 285

Lys Phe Tyr Ile Phe Arg Ser Asp Glu Ser Ala Asn Leu Thr Gln Gly

290

295

300

Phe Met Tyr Phe Asn Val Asn Gln Thr Ile Gln Glu Val Glu Asn Val
305 310 315 320

Asp Leu Ser Glu Phe Met Arg Arg Ile Ser Gly Ser Glu Ile Val Phe
325 330 335

Leu Phe Ser Leu Phe Gly Phe Val Trp Leu Leu Arg Lys His Lys Ser
340 345 350

Met Ile Met Ala Leu Pro Ile Leu Val Leu Gly Phe Leu Ala Leu Lys
355 360 365

Gly Gly Leu Arg Phe Thr Ile Tyr Ser Val Pro Val Met Ala Leu Gly
370 375 380

[0815]

Phe Gly Phe Leu Leu Ser Glu Phe Lys Ala Ile Met Val Lys Lys Tyr
385 390 395 400

Ser Gln Leu Thr Ser Asn Val Cys Ile Val Phe Ala Thr Ile Leu Thr
405 410 415

Leu Ala Pro Val Phe Ile His Ile Tyr Asn Tyr Lys Ala Pro Thr Val
420 425 430

Phe Ser Gln Asn Glu Ala Ser Leu Leu Asn Gln Leu Lys Asn Ile Ala
435 440 445

Asn Arg Glu Asp Tyr Val Val Thr Trp Ala Ala Tyr Gly Tyr Pro Val
450 455 460

Arg Tyr Tyr Ser Asp Val Lys Thr Leu Val Asp Gly Gly Lys His Leu
465 470 475 480

Gly Lys Asp Asn Phe Phe Pro Ser Phe Ala Leu Ser Lys Asp Glu Gln
485 490 495

Ala Ala Ala Asn Met Ala Arg Leu Ser Val Glu Tyr Thr Glu Lys Ser
500 505 510

Phe Tyr Ala Pro Gln Asn Asp Ile Leu Lys Thr Asp Ile Leu Gln Ala
515 520 525

Met Met Lys Asp Tyr Asn Gln Ser Asn Val Asp Leu Phe Leu Ala Ser
530 535 540

Leu Ser Lys Pro Asp Phe Lys Ile Asp Thr Pro Lys Thr Arg Asp Ile
545 550 555 560

[0816] Tyr Leu Tyr Met Pro Ala Arg Met Ser Leu Ile Phe Ser Thr Val Ala
565 570 575

Ser Phe Ser Phe Ile Asn Leu Asp Thr Gly Val Leu Asp Lys Pro Phe
580 585 590

Thr Phe Ser Thr Ala Tyr Pro Leu Asp Val Lys Asn Gly Glu Ile Tyr
595 600 605

Leu Ser Asn Gly Val Val Leu Ser Asp Asp Phe Arg Ser Phe Lys Ile
610 615 620

Gly Asp Asn Val Val Ser Val Asn Ser Ile Val Glu Ile Asn Ser Ile
625 630 635 640

Lys Gln Gly Glu Tyr Lys Ile Thr Pro Ile Asp Asp Lys Ala Gln Phe
645 650 655

Tyr Ile Phe Tyr Leu Lys Asp Ser Ala Ile Pro Tyr Ala Gln Phe Ile
660 665 670

Leu Met Asp Lys Thr Met Phe Asn Ser Ala Tyr Val Gln Met Phe Phe
675 680 685

[0817] Leu Gly Asn Tyr Asp Lys Asn Leu Phe Asp Leu Val Ile Asn Ser Arg
690 695 700

Asp Ala Lys Val Phe Lys Leu Lys Ile Tyr Pro Tyr Asp Val Pro Asp
705 710 715 720

Tyr Ala

[0818] SEQ ID NO:27-pCC1FOS空质粒的核苷酸序列

基因座 具有 MCS 盒的 pCC1FOS

特征 位置/合格者

区域 230..256

/标记物="pCC1/pEpiFOS fwd"

区域 311..330

/标记物="T7 启动子"

区域 补体(504..529)

/标记物="pCC1pEpiFOS rv"

[0819]

CDS 补体(805..1464)

/标记物=cat

CDS 1683..2030

/标记物=redF

CDS 3425..4180

/标记物=repE

CDS 4759..5934

/标记物=parA

CDS 5934..6905

[0820]

/标记物=parB

[0821] ORIGIN

[0822] 长度:8171bp

[0823] 类型:DNA 环状UNA

[0824] 生物体:人工的

[0825] 序列:

[0826] 1 GCGGCCGCAA GGGGTTTCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
[0827] 61 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT
[0828] 121 GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGCTGCGCA ACTGTTGGGA
[0829] 181 AGGGCGATCG GTGCGGGCCT CTTCGCTATT ACGCCAGCTG GCGAAAGGGG GATGTGCTGC
[0830] 241 AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCAGTCA CGACGTTGTA AAACGACGGC
[0831] 301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG CTCGGTACCC GGGGATCCCA
[0832] 361 CGTGGCGCGC CACTAGTGCT AGCGACGTCG TGGGATCCTC TAGAGTCGAC CTGCAGGCAT
[0833] 421 GCAAGCTTGA GTATTCTATA GTCTCACCTA AATAGCTTGG CGTAATCATG GTCATAGCTG
[0834] 481 TTTCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA
[0835] 541 AAGTGTAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAAATTGC GTTGCGCTCA
[0836] 601 CTGCCCCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC
[0837] 661 GAACCCCTTG CGGCCGCCG GGCCGTCGAC CAATTCTCAT GTTTGACAGC TTATCATCGA
[0838] 721 ATTTCTGCCA TTCATCCGCT TATTATCACT TATTCAGGCG TAGCAACCAG GCGTTTAAGG
[0839] 781 GCACCAATAA CTGCCTTAAA AAAATTACGC CCCGCCCTGC CACTCATCGC AGTACTGTTG
[0840] 841 TAATTCATTA AGCATTCTGC CGACATGGAA GCCATCACAA ACGGCATGAT GAACCTGAAT
[0841] 901 CGCCAGCGGC ATCAGCACCT TGTCGCCTTG CGTATAATAT TTGCCCATGG TGAAAACGGG
[0842] 961 GGCGAAGAAG TTGTCCATAT TGGCCACGTT TAAATCAAAA CTGGTGAAAC TCACCCAGGG
[0843] 1021 ATTGGCTGAG ACGAAAAACA TATTCTCAAT AAACCCTTTA GGGAAATAGG CCAGGTTTTC
[0844] 1081 ACCGTAACAC GCCACATCTT GCGAATATAT GTGTAGAAAC TGCCGGAAAT CGTCGTGGTA
[0845] 1141 TTCACTCCAG AGCGATGAAA ACGTTTCAGT TTGCTCATGG AAAACGGTGT AACAAGGGTG
[0846] 1201 AACACTATCC CATATCACCA GCTCACCGTC TTTCATTGCC ATACGAAATT CCGGATGAGC
[0847] 1261 ATTCATCAGG CGGGCAAGAA TGTGAATAAA GGCCGGATAA AACTTGTGCT TATTTTTCTT
[0848] 1321 TACGGTCTTT AAAAAGGCCG TAATATCCAG CTGAACGGTC TGGTTATAGG TACATTGAGC
[0849] 1381 AACTGACTGA AATGCCTCAA AATGTTCTTT ACGATGCCAT TGGGATATAT CAACGGTGGT
[0850] 1441 ATATCCAGTG ATTTTTTTCT CCATTTTAGC TTCCTTAGCT CCTGAAAATC TCGATAACTC
[0851] 1501 AAAAAATACG CCCGGTAGTG ATCTTATTTT ATTATGGTGA AAGTTGGAAC CTCTTACGTG
[0852] 1561 CCGATCAACG TCTCATTTTC GCCAAAAGTT GGCCCGGGC TTCCCGGTAT CAACAGGGAC
[0853] 1621 ACCAGGATTT ATTTATTCTG CGAAGTGATC TTCCGTCACA GGTATTTATT CGCGATAAGC
[0854] 1681 TCATGGAGCG GCGTAACCGT CGCACAGGAA GGACAGAGAA AGCGCGGATC TGGGAAGTGA
[0855] 1741 CGGACAGAAC GGTGAGGACC TGGATTGGGG AGGCGGTTGC CGCCGCTGCT GCTGACGGTG
[0856] 1801 TGACGTTCTC TGTTCGGTC ACACCACATA CGTTCGCCA TTCCTATGCG ATGCACATGC
[0857] 1861 TGTATGCCGG TATACCGCTG AAAGTTCTGC AAAGCCTGAT GGGACATAAG TCCATCAGTT
[0858] 1921 CAACGGAAGT CTACACGAAG GTTTTTGCGC TGGATGTGGC TGCCCGGCAC CGGGTGCAGT
[0859] 1981 TTGCGATGCC GGAGTCTGAT GCGGTTGCGA TGCTGAAACA ATTATCCTGA GAATAAATGC
[0860] 2041 CTTGGCCTTT ATATGGAAAT GTGGAAGTGA GTGGATATGC TGTTTTTGTC TGTTAAACAG
[0861] 2101 AGAAGCTGGC TGTATCCAC TGAGAAGCGA ACGAAACAGT CGGGAAAATC TCCATTATC

[0862] 2161 GTAGAGATCC GCATTATTAA TCTCAGGAGC CTGTGTAGCG TTTATAGGAA GTAGTGTTC
[0863] 2221 GTCATGATGC CTGCAAGCGG TAACGAAAAC GATTTGAATA TGCCCTCAGG AACAAATAGAA
[0864] 2281 ATCTTCGTGC GGTGTTACGT TGAAGTGGAG CGGATTATGT CAGCAATGGA CAGAACAACC
[0865] 2341 TAATGAACAC AGAACCATGA TGTGGTCTGT CCTTTTACAG CCAGTAGTGC TCGCCGCAGT
[0866] 2401 CGAGCGACAG GGC GAAGCCC TCGGCTGGTT GCCCTCGCCG CTGGGCTGGC GGCCGTCTAT
[0867] 2461 GGCCCTGCAA ACGCGCCAGA AACGCCGTGC AAGCCGTGTG CGAGACACCG CGGCCGGCCG
[0868] 2521 CCGGCGTTGT GGATACCTCG CGGAAAACCTT GGCCCTCACT GACAGATGAG GGGCGGACGT
[0869] 2581 TGACACTTGA GGGGCCGACT CACCCGGCGC GGCGTTGACA GATGAGGGGC AGGCTCGATT
[0870] 2641 TCGGCCGGCG ACGTGGAGCT GGCCAGCCTC GCAAATCGGC GAAAACGCCT GATTTTACGC
[0871] 2701 GAGTTTCCCA CAGATGATGT GGACAAGCCT GGGGATAAGT GCCCTGCGGT ATTGACACTT
[0872] 2761 GAGGGGCGCG ACTACTGACA GATGAGGGGC GCGATCCTTG ACACTTGAGG GGCAGAGTGC
[0873] 2821 TGACAGATGA GGGGCGCACC TATTGACATT TGAGGGGCTG TCCACAGGCA GAAAATCCAG
[0874] 2881 CATTTGCAAG GGTTTCGCC CGTTTTTCGG CCACCGCTAA CCTGTCTTTT AACCTGCTTT
[0875] 2941 TAAACCAATA TTTATAAACC TTGTTTTTAA CCAGGGCTGC GCCCTGTGCG CGTGACCGCG
[0876] 3001 CACGCCGAAG GGGGGTGCCC CCCCTTCTCG AACCCCTCCG GTCGAGTGAG CGAGGAAGCA
[0877] 3061 CCAGGGAACA GCACTTATAT ATTCTGCTTA CACACGATGC CTGAAAAAAC TTCCCTTGGG
[0878] 3121 GTTATCCACT TATCCACGGG GATATTTTAA TAATTATTTT TTTTATAGTT TTTAGATCTT
[0879] 3181 CTTTTTTAGA GCGCCTTGTA GGCCTTTATC CATGCTGGTT CTAGAGAAGG TGTTGTGACA
[0880] 3241 AATTGCCCTT TCAGTGTGAC AAATCACCTT CAAATGACAG TCCTGTCTGT GACAAATTGC
[0881] 3301 CCTTAACCCT GTGACAAATT GCCCTCAGAA GAAGCTGTTT TTTCACAAAG TTATCCCTGC
[0882] 3361 TTATTGACTC TTTTTTATTT AGTGTGACAA TCTAAAAACT TGTCACTT CACATGGATC
[0883] 3421 TGTCATGGCG GAAACAGCGG TTATCAATCA CAAGAAACGT AAAAATAGCC CGCGAATCGT
[0884] 3481 CCAGTCAAAC GACCTCACTG AGGCGGCATA TAGTCTCTCC CGGGATCAAA AACGTATGCT
[0885] 3541 GTATCTGTTT GTTGACCAGA TCAGAAAATC TGATGGCACC CTACAGGAAC ATGACGGTAT
[0886] 3601 CTGCGAGATC CATGTTGCTA AATATGCTGA AATATTCGGA TTGACCTCTG CGGAAGCCAG
[0887] 3661 TAAGGATATA CGGCAGGCAT TGAAGAGTTT CGCGGGGAAG GAAGTGGTTT TTTATCGCCC
[0888] 3721 TGAAGAGGAT GCCGGCGATG AAAAAGGCTA TGAATCTTTT CCTTGGTTTA TCAAACGTGC
[0889] 3781 GCACAGTCCA TCCAGAGGGC TTTACAGTGT ACATATCAAC CCATATCTCA TTCCCTTCTT
[0890] 3841 TATCGGGTTA CAGAACCGGT TTACGCAGTT TCGGCTTAGT GAAACAAAAG AAATCACCAA
[0891] 3901 TCCGTATGCC ATGCGTTTAT ACGAATCCCT GTGTCAGTAT CGTAAGCCGG ATGGCTCAGG
[0892] 3961 CATCGTCTCT CTGAAAATCG ACTGGATCAT AGAGCGTTAC CAGCTGCCTC AAAGTTACCA
[0893] 4021 GCGTATGCCT GACTTCCGCC GCCGCTTCTT GCAGGTCTGT GTTAATGAGA TCAACAGCAG
[0894] 4081 AACTCCAATG CGCCTCTCAT ACATTGAGAA AAAGAAAGGC CGCCAGACGA CTCATATCGT
[0895] 4141 ATTTTCCTTC CGGATATCA CTTCCATGAC GACAGGATAG TCTGAGGGTT ATCTGTCACA
[0896] 4201 GATTTGAGGG TGGTTCGTCA CATTTGTTCT GACCTACTGA GGGTAATTTG TCACAGTTTT
[0897] 4261 GCTGTTTCCT TCAGCCTGCA TGGATTTTCT CATACTTTT GAACTGTAAT TTTTAAGGAA
[0898] 4321 GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCCCTTCT CTTTCCCTTC GTCATGTGAC
[0899] 4381 CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG ATTATCACAG TTTATTACTC
[0900] 4441 TGAATTGGCT ATCCGCGTGT GTACCTCTAC CTGGAGTTTT TCCCACGGTG GATATTTCTT

[0901] 4501 CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT CTTTGCTTCC TCGCCAGTTC
[0902] 4561 GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT GATAATAAGT GACTGAGGTA
[0903] 4621 TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT TAAACAACCT TGCGGTTTTT
[0904] 4681 TGATGACTTT GCGATTTTGT TGTTGCTTTG CAGTAAATTG CAAGATTTAA TAAAAAACG
[0905] 4741 CAAAGCAATG ATTAAAGGAT GTTCAGAATG AAACCTCATGG AAACACTTAA CCAGTGCATA
[0906] 4801 AACGCTGGTC ATGAAATGAC GAAGGCTATC GCCATTGCAC AGTTTAATGA TGACAGCCCC
[0907] 4861 GAAGCGAGGA AAATAACCCG GCGCTGGAGA ATAGGTGAAG CAGCGGATTT AGTTGGGGTT
[0908] 4921 TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC TACCGCACCC GGATATGGAA
[0909] 4981 ATTCGAGGAC GGGTTGAGCA ACGTGTTGGT TATACAATTG AACAAATTAA TCATATGCGT
[0910] 5041 GATGTGTTTG GTACGCGATT GCGACGTGCT GAAGACGTAT TTCCACCGGT GATCGGGGTT
[0911] 5101 GCTGCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG TTCATCTTGC TCAGGATCTG
[0912] 5161 GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG ACCCCCAGGG AACAGCCTCA
[0913] 5221 ATGTATCACG GATGGGTACC AGATCTTCAT ATTCATGCAG AAGACACTCT CCTGCCTTTC
[0914] 5281 TATCTTGGGG AAAAGGACGA TGTCACCTAT GCAATAAAGC CCACTTGCTG GCCGGGGCTT
[0915] 5341 GACATTATTC CTTCTGTCT GGCTCTGCAC CGTATTGAAA CTGAGTTAAT GGGCAAATTT
[0916] 5401 GATGAAGGTA AACTGCCCAC CGATCCACAC CTGATGCTCC GACTGGCCAT TGAAACTGTT
[0917] 5461 GCTCATGACT ATGATGTCAT AGTTATTGAC AGCGCGCCTA ACCTGGGTAT CGGCACGATT
[0918] 5521 AATGTCGTAT GTGCTGCTGA TGTGCTGATT GTTCCCACGC CTGCTGAGTT GTTTGACTAC
[0919] 5581 ACCTCCGCAC TGCAGTTTTT CGATATGCTT CGTGATCTGC TCAAGAACGT TGATCTTAAA
[0920] 5641 GGGTTCGAGC CTGATGTACG TATTTTGCTT ACCAAATACA GCAATAGTAA TGGCTCTCAG
[0921] 5701 TCCCCGTGGA TGGAGGAGCA AATTCCGGAT GCCTGGGGAA GCATGGTTCT AAAAAATGTT
[0922] 5761 GTACGTGAAA CGGATGAAGT TGGTAAAGGT CAGATCCGGA TGAGAACTGT TTTTGAACAG
[0923] 5821 GCCATTGATC AACGCTCTTC AACTGGTGCC TGGAGAAATG CTCCTTCTAT TTGGGAACCT
[0924] 5881 GTCTGCAATG AAATTTTCGA TCGTCTGATT AAACCACGCT GGGAGATTAG ATAATGAAGC
[0925] 5941 GTGCGCCTGT TATTCCAAAA CACACGCTCA ATACTCAACC GGTTGAAGAT ACTTCGTTAT
[0926] 6001 CGACACCAGC TGCCCCGATG GTGGATTGCT TAATTGCGCG CGTAGGAGTA ATGGCTCGCG
[0927] 6061 GTAATGCCAT TACTTTGCCT GTATGTGGTC GGGATGTGAA GTTTACTCTT GAAGTGCTCC
[0928] 6121 GGGGTGATAG TGTTGAGAAG ACCTCTCGGG TATGGTCAGG TAATGAACGT GACCAGGAGC
[0929] 6181 TGCTTACTGA GGACGCACTG GATGATCTCA TCCCTTCTTT TCTACTGACT GGTCAACAGA
[0930] 6241 CACCGGCGTT CGGTCGAAGA GTATCTGGTG TCATAGAAAT TGCCGATGGG AGTCGCCGTC
[0931] 6301 GTAAAGCTGC TGCACTTACC GAAAGTGATT ATCGTGTCTT GGTTGGCGAG CTGGATGATG
[0932] 6361 AGCAGATGGC TGCATTATCC AGATTGGGTA ACGATTATCG CCCAACAAGT GCTTATGAAC
[0933] 6421 GTGGTCAGCG TTATGCAAGC CGATTGCAGA ATGAATTTGC TGGAAATATT TCTGCGCTGG
[0934] 6481 CTGATGCGGA AAATATTTCA CGTAAGATTA TTACCCGCTG TATCAACACC GCCAAATTGC
[0935] 6541 CTAAATCAGT TGTTGCTCTT TTTTCTCACC CCGGTGAACT ATCTGCCCCG TCAGGTGATG
[0936] 6601 CACTTCAAAA AGCCTTTACA GATAAAGAGG AATTACTTAA GCAGCAGGCA TCTAACCTTC
[0937] 6661 ATGAGCAGAA AAAAGCTGGG GTGATATTTG AAGCTGAAGA AGTTATCACT CTTTTAACTT
[0938] 6721 CTGTGCTTAA AACGTCATCT GCATCAAGAA CTAGTTTAAAG CTCACGACAT CAGTTTGCTC
[0939] 6781 CTGGAGCGAC AGTATTGTAT AAGGGCGATA AAATGGTGCT TAACCTGGAC AGGTCTCGTG

- [0940] 6841 TTCCAAC TGA GTGTATAGAG AAAATTGAGG CCATTCTTAA GGAAC TTGAA AAGCCAGCAC
- [0941] 6901 CCTGATGCGA CCACGTTT TA GTCTACGTTT ATCTGTCTTT ACTTAATGTC CTTTGT TACA
- [0942] 6961 GGCCAGAAAG CATAACTGGC CTGAATATTC TCTCTGGGCC CACTGTTCCA CTTGTATCGT
- [0943] 7021 CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT CGTCGGTCTG ATTATTAGTC
- [0944] 7081 TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA GTCTGGGACC ACGGTCCCAC
- [0945] 7141 TCGTATCGTC GGTCTGATAA TCAGACTGGG ACCACGGTCC CACTCGTATC GTCGGTCTGA
- [0946] 7201 TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC TGATTATTAG TCTGGGACCA
- [0947] 7261 CGGTCCC ACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA CCACGGTCCC ACTCGTATCG
- [0948] 7321 TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA TCGTCGGTCT GATTATTAGT
- [0949] 7381 CTGGGACCAC GATCCC ACTC GTGTTGTCGG TCTGATTATC GGTCTGGGAC CACGGTCCCA
- [0950] 7441 CTTGTATTGT CGATCAGACT ATCAGCGTGA GACTACGATT CCATCAATGC CTGTCAAGGG
- [0951] 7501 CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA CCTCGGTGTG CGGTTGTATG
- [0952] 7561 CCTGCTGTGG ATTGCTGCTG TGTCCTGCTT ATCCACAACA TTTTGCGCAC GGTTATGTGG
- [0953] 7621 ACAAATACC TGGTTACCCA GGCCGTGCCG GCACGTAAAC CGGGCTGCAT CCGATGCAAG
- [0954] 7681 TGTGTCGCTG TCGACGAGCT CGCGAGCTCG GACATGAGGT TGCCCCGTAT TCAGTGTGCG
- [0955] 7741 TGATTTGTAT TGTCTGAAGT TGTTTTTACG TTAAGTTGAT GCAGATCAAT TAATACGATA
- [0956] 7801 CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT CCACGCACGT TGTGATATGT
- [0957] 7861 AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA TCCGACAGGT TACGGGGCGG
- [0958] 7921 CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT TTAAGGCGTT TCCGTTCTTC
- [0959] 7981 TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA AAAGAAAGGA AACGACAGGT
- [0960] 8041 GCTGAAAGCG AGCTTTTTGG CCTCTGTCGT TTCCTTTCTC TGTTTTTGTC CGTGGAATGA
- [0961] 8101 ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG CATACATTAT ACGAAGTTAT
- [0962] 8161 ATTCGATCCA C
- [0963] SEQ ID NO:28-pCC1FOS cut(pFOS)和弗氏志贺氏菌(*S.flexneri*)60-抗原无
- [0964] Z3206的核苷酸序列
- [0965] 基因座 pFOS cut和0-抗原cut(-Z3206)
- [0966] 定义具有MCS盒并用NheI和AscI切割的反向pCC1FOS连接到用galFNheI
- [0967] 和wzzAscI扩增并用NheI和AscI切割的弗氏志贺氏菌(*S.flexneri*)60抗原簇

特征	位置/合格者
[0968]	CDS 3..411 /标记物=galF
	CDS 784..1869 /标记物=rmlB
	CDS 1869..2768 /标记物=rmlD
	CDS 2826..3704 /标记物=rmlA
	CDS 3709..4266 /标记物=rmlC
	CDS 4263..5495 /标记物=wzx
	CDS 5551..6738

		/标记物=wzy
	CDS	6755..7624
		/标记物=wfbY
	CDS	7621..8454
		/标记物=wfbZ
	CDS	8559..9965
		/标记物=gnd
	CDS	10187..11380
		/标记物=ugd
	CDS	补体(11446..12450)
		/标记物=uqe
	CDS	12802..12828
		/标记物=wzz'
	区域	补体(12868..12887)
[0969]		/标记物="T7 启动子"
	区域	补体(12942..12968)
		/标记物="pCC1/pEpiFOS fwd"
	CDS	补体(14460..15431)
		/标记物=parB
	CDS	补体(15431..16606)
		/标记物=parA
	CDS	补体(17185..17940)
		/标记物=repE
	CDS	补体(19335..19682)
		/标记物=redF
	CDS	19901..20560
		/标记物=cat
	区域	20836..20861
		/标记物="pCC1pEpiFOS rv"
[0970]	长度:20982bp	
[0971]	类型:DNA 环状UNA	

[0972] 序列:

[0973] 1 CTAGCGGCAA AACGTATGCC GGGTGACCTC TCTGAATACT CCGTCATCCA GACCAAAGAA
[0974] 61 CCGCTGGATC GCGAAGGTAA AGTCAGCCGC ATTGTTGAAT TTATCGAAAA ACCGGATCAG
[0975] 121 CCGCAGACGC TGGACTCAGA CATCATGGCC GTTGGTCGCT ATGTGCTTTC TGCCGATATT
[0976] 181 TGGCCGGAAC TTGAACGTAC TCAGCCTGGT GCATGGGGAC GTATTCAGCT GACTGATGCC
[0977] 241 ATTGCCGAGC TGGCGAAAAA ACAGTCCGTT GATGCAATGC TGATGACCGG CGACAGCTAC
[0978] 301 GACTGCGGTA AAAAAATGGG CTATATGCAG GCGTTTGTGA AGTATGGGCT GCGCAACCTG
[0979] 361 AAAGAAGGGG CGAAGTCCG TAAAGGTATT GAGAAGCTGT TAAGCGAATA ATGAAAATCT
[0980] 421 GACCGGATGT AACGGTTGAT AAGAAAATTA TAACGGCAGT GAAGATTCGT GGTGAAAGTA
[0981] 481 ATTTGTTGCG AATATTCCTG CCGTTGTTTT ATATAAACAA TCAGAATAAC AACGAGTTAG
[0982] 541 CAATAGGATT TTAGTCAAAAG TTTTCCAGGA TTTTCCTTGT TTCCAGAGCG GATTGGTAAG
[0983] 601 ACAATTAGCT TTTGAATTTT TCGGGTTTAG CGCGAGTGGG TAACGCTCGT CACATCGTAG
[0984] 661 GCATGCATGC AGTGCTCTGG TAGCTGTAAA GCCAGGGGCG GTAGCGTGCA TTAATACTTC
[0985] 721 TATTAATCAA ACTGAGAGCC GCTTATTTCA CAGCATGCTC TGAAGCAATA TGAATAAAT
[0986] 781 TAGGTGAAAA TACTTGTTAC TGGTGGCGCA GGATTTATTG GTTTTGCTGT AGTTCGTCAC
[0987] 841 ATTATAAATA ATACGCAGGA TAGTGTTGTT AATGTCGATA AATTAACGTA CGCCGGAAC
[0988] 901 CTGGAATCAC TTGCTGATGT TTCTGATTCT GAACGCTATG TTTTGAACA TGCGGATATT
[0989] 961 TGCGATGCAG CTGCAATGGC ACGGATTTTT GCTCAGCATC AGCCAGATGC AGTGATGCAC
[0990] 1021 CTGGCTGCTG AAAGCCATGT TGACCGTTCA ATTACAGGTC CTGCGGCATT TATTGAAACC
[0991] 1081 AATATTGTTG GTACATATGT CCTTTTGGA GCCGCTCGCA ATTATTGGTC TGCTCTTGAT
[0992] 1141 AGCGACAAGA AAAGTAGATT CCGTTTTTCAT CATATTTCTA CTGACGAAGT CTATGGTGAT
[0993] 1201 TTGCCTCATC CTGACGAGGT AAATAATACA GAAGAATTAC CCTTATTTAC AGAGACAACA
[0994] 1261 GCTTACGCGC CAAGCAGCCC TTATTCCGCT TCAAAAGCAT CCAGCGATCA TTTAGTCCGC
[0995] 1321 GCGTGGAAC GTACCTATGG TTTACCAACC ATTGTGACTA ATTGCTCTAA TAATTATGGT
[0996] 1381 CCTTATCATT TCCCGAAAAA ATTGATTCCA TTGGTTATTC TGAATGCTCT GGAAGGTAAG
[0997] 1441 GCATTACCTA TTTATGGCAA AGGGGATCAA ATTCGTGACT GGCTGTATGT TGAAGATCAT
[0998] 1501 GCGCGTGCGT TATATACCGT CGTAACCGAA GGTAAGCGG GTGAACTTA TAACATTGGT
[0999] 1561 GGACACAACG AAAAGAAAAA CATCGATGTA GTGCTCACTA TTTGTGATTT GCTGGATGAG
[1000] 1621 ATTGTACCGA AAGAGAAATC TTACCGCGAG CAAATTACTT ATGTTGCCGA TCGCCCGGGA
[1001] 1681 CACGATCGCC GTTATGCGAT TGATGCAGAG AAGATTAGCC GCGAATTGGG CTGGAAACCG
[1002] 1741 CAGGAAACGT TTGAGAGCGG GATTCGGAAG ACATTGGAAT GGTACCTGTC CAATACAAAA
[1003] 1801 TGGGTTGATA ATGTGAAAAG TGGTGCTTAT CAATCGTGGA TTGAACAGAA CTATGAGGGC
[1004] 1861 CGCCAGTAAT GAATATCCTC CTTTTCGGCA AAACAGGGCA GGTAGGTTGG GAACTACAGC
[1005] 1921 GTGCTCTGGC ACCTTTGGGT AATTTGATTG CTCTTGATGT TCACTCCACT GATTATTGTG
[1006] 1981 GTGATTTTAG TAATCCTGAA GGTGTAGCTG AAACAGTCAA AAGAATTCGA CCTGATGTTA
[1007] 2041 TTGTTAATGC TGGCGCTCAC ACCGCGTAG ATAAGGCTGA GTCAGAACCC GAATTTGCAC
[1008] 2101 AATTACTCAA TGCGACTAGT GTTGAATCAA TTGCAAAAGA GGCTAATGAA GTTGGGGCTT
[1009] 2161 GGGTAATTCA TTAACAATCT GACTACGTAT TCCCTGGAAG TGGCGACACG CCATGGCTGG
[1010] 2221 AGACGGATGC AACCGCACCG CTAAATGTTT ACGGTGAAAC CAAGTTAGCC GGAGAAAAAG

[1011] 2281 CGTTACAGGA ACATTGCGCG AAGCATCTTA TTTCCGTAC CAGCTGGGTA TACGCAGCTA
[1012] 2341 AAGGAAATAA CTTCCGCAAA ACGATGTTGC GTCTGGCAAA AGAGCGCGAA GAACTGGCTG
[1013] 2401 TGATAAATGA TCAATTTGGT GCGCCAACAG GTGCTGAGCT GCTGGCTGAT TGTACGGCAC
[1014] 2461 ATGCTATTCG TGTGGCACTG AATAAACCGG AAGTCGCAGG TTTGTACCAT CTGGTAGCCA
[1015] 2521 GTGGTACCAC AACCTGGCAC GATTATGCTG CGCTGGTTTT TGAAGAGGCG CGCAAAGCAG
[1016] 2581 GTATTCCTCT TGCCTCAAC AAGCTCAACG CAGTACCAAC AACAGCCTAT CCTACACCAG
[1017] 2641 CTCGTCGTCC ACATAACTCT CGCCTTAATA CAGAAAAATT TCAGCAGAAC TTTGCGCTTG
[1018] 2701 TCTTGCTGA CTGGCAGGTT GGTGTGAAAC GAATGCTCAA CGAATTAATT ACGACTACAG
[1019] 2761 CAATTTAATA GTTTTTGCAT CTTGTTCTGT ATGGTGGAGC AAGATGAATT AAAAGGAATG
[1020] 2821 ATGAAATGAA AACGCGTAAA GGTATTATTT TAGCGGGTGG TTCTGGTACA CGTCTTTATC
[1021] 2881 CTGTGACTAT GGCTGTCAGT AAACAGCTAT TACCTATTTA TGATAAGCCG ATGATCTATT
[1022] 2941 ACCCGCTCTC TACACTGATG TTGGCGGGTA TTCGCGATAT TCTGATTATT AGTACGCCAC
[1023] 3001 AGGATACTCC TCGTTTTCAA CAACTGCTAG GTGACGGTAG CCAGTGGGGG CTAAATCTTC
[1024] 3061 AGTACAAAGT GCAACCGACT CCAGATGGGC TTGCGCAGGC GTTTATTATC GGTGAAGAGT
[1025] 3121 TTATCGGTGG TGATGATTGT GCTTTGGTTC TTGGTGATAA TATCTTCTAC GGTGATGATC
[1026] 3181 TGCCGAAGTT AATGGATGTC GCTGTTAACA AAGAAAGTGG TGCAACGGTA TTTGCCTATC
[1027] 3241 ACGTTAATGA TCCTGAACGC TACGGCGTCG TTGAGTTTGA TAAAAACGGT ACGGCAATAA
[1028] 3301 GCCTGGAAGA AAAACCGCTA CAACCAAAAA GTAATTATGC GGTAACCGGG CTTTATTTCT
[1029] 3361 ATGATAACGA CGTTGTGCGA ATGGCGAAAA ACCTTAAGCC TTCTGCCCGT GGTGAACTGG
[1030] 3421 AAATTACCGA TATTAACCGT ATTTATATGG AACAGGGGCG TTTATCCGTT GCCATGATGG
[1031] 3481 GGCGTGGTTA TGCATGGCTG GATACGGGGA CACATCAGAG TCTTATTGAA GCAAGCAACT
[1032] 3541 TCATTGCCAC CATTGAAGAG CGCCAGGGAC TAAAGGTTTC CTGCCCAGAA GAAATTGCTT
[1033] 3601 ACCGTAAAGG GTTTATTGAT GCTGAACAGG TGAAAGCATT AGCGGAGCCG CTGAAAAAAA
[1034] 3661 ATGCTTATGG ACAGTATCTG CTGAAAATGA TTAAAGGTTA TTAATAAAAT GAACGTAATT
[1035] 3721 AAAACAGAAA TTCCTGATGT GTTAATTTTC GAGCCGAAAG TTTTGGTGA TGAGCGTGGT
[1036] 3781 TTCTTTATGG AAAGCTTTAA TCAGAAAGTT TTCGAAGAAG CTGTAGGACG TAAGGTTGAA
[1037] 3841 TTTGTTTCAGG ATAACCATTC GAAGTCTAGT AAAGGTGTTT TACGCGGGCT GCATTATCAG
[1038] 3901 TTAGAACCTT ATGCGCAAGG GAAACTGGTA CGTTGCGTTG TTGGTGAGGT TTTTGATGTA
[1039] 3961 GCTGTTGATA TTCGTAAATC GTCGCCTACC TTTGGTAAAT GGGTTGGGGT GAATTTATCT
[1040] 4021 GCTGAGAATA AGCGGCAATT GTGGATCCCT GAGGGATTG CACATGGTTT TTTGGTGCTG
[1041] 4081 AGCGAGACTG CGGAATTTTT ATATAAAACG ACGAACTATT ATCATCCTGA TAGTGATAGA
[1042] 4141 GGGATTGTAT GGAATGATCC TATTCTGAGC ATAAAATGGC CGACGATAGA ACATAATAAT
[1043] 4201 TATATTTTAT CGATTAAAGA TGCAAGGGCT AAAGAATTGC ATAACATGAA GGAATTATTT
[1044] 4261 TTGTGAGTAT TGTAAGAAT ACTTTATGGA ATATAAGTGG GTATATTATA CCATCATTA
[1045] 4321 TAGCAATTCC TGCGTTAGGT ATACTGTCTA GAATTCCTAGG GACCGAGCAA TTTGGCCTTT
[1046] 4381 TTACGTTAGC TATTGCCTTA GTTGATATG CAAGTATTTT TGATGCTGGA TTGACCAGAG
[1047] 4441 CTGTTATAAG AGAAGTATCA ATATATAAAA ATGTTTCATA AGAATTAAGA GCGATCATTT
[1048] 4501 CAACTTCAAC GGTAAATTCTA ACTATATTGG GCTTGATTGG CGGTAGTGTA CTATTTTTGA
[1049] 4561 GTAGCAATGT AATTGTTAAA TTATTAAACA TTAACGCGAA TCATGTTGTA GAATCTGTCA

[1050] 4621 AAGCAATATA TATTATTTCA GCTACCATAC CCTTATACTT GTTAAACCAA GTCTGGTTGG
[1051] 4681 GGATTTTTGA GGGGATGGAA AAGTTCAGAA AAGTAAATTT AATAAAATCA ATTAACAACCT
[1052] 4741 CTTTTGTGGC TGGATTACCA GTGATTTTCT GTTTTTTTCA TGGAGGATTA CTAAGTGCTA
[1053] 4801 TATATGTTTT AGTTATGGCA AGAGTCTTAT CACTTATAGT GACCTTTATA TTTAGTCGAA
[1054] 4861 AACTAATAAT ATCATCTGGG CTGTCTGTAA AAATTGTAAC AGTTAAAAGA TTAATCGGCT
[1055] 4921 TTGGAAGCTG GATAACAGTT AGCAATATTA TTAGCCCTAT TATGACATAT ATGGATCGTT
[1056] 4981 TTATTCCTTC ACACATTGTG GGGGCTGATA AAGTTTCTTT TTATACTGCT CCGTCTGAAG
[1057] 5041 GTATACAACG CTTAACGATA TTACCAAGTG CGTTGTCCAG AGCTATTTTT CCAAGATTAA
[1058] 5101 GTTCAGAATT GCAATCGGTA AAGCAAACCTA AAATATTATC ATATTTTATA ATGGTTATTG
[1059] 5161 GTATACTTCC AATTGTAATG TTGATAATTA TTTTATCAGA TTTTATAATG TCCGCTTGGA
[1060] 5221 TGGGACCTAC ATATCATGGG ACGCCAGGTA TAGTATTTAA AATTCTTGCA ATAGGTTTCT
[1061] 5281 TTTTAAATTG CATTGCACAA ATCCCATTTG TTTCAGTTCA GGCTAGTGGA AGATCAAAAA
[1062] 5341 TTACAGCTAT TATTCATTTG CTCGAAGTTA TCCCATATTT ATGCATATTA TATATTTTTA
[1063] 5401 TTTATCATTG GGAATTGTT GGAGCCGCAA TAGCATGGTC TGTAAGAACA TCGTTAGATT
[1064] 5461 TTTTGATATT ATTATTAATT GATACGAAAT ATTAATAGCG AATTGATTTT AGGGATTACT
[1065] 5521 TCCTCAAGCC CATCTAATTA GAGTGCAAAC ATGACTTCTG ATTTTTATAA CTCAAAAGAC
[1066] 5581 AAAAGTTTAA GTGTTCTTTT GTTTTTTGGG TTTATATTTT TCCTTACACG TAGCTTTCCA
[1067] 5641 TTTATTCAAT ATAGTTGGAT TATGGAGGGG TTTTATGTC TTTGTATCAT GTCATTTACA
[1068] 5701 AAGAAAATTG CAAACGGAAT ATATCACTAT CCTGTTATTT TAATATTTCT ATTAGCTCTT
[1069] 5761 TTTATAAATT TTATTTATTC CTATATCAAG GGTAACGATA TAGCGATAAT AATTAGGTTT
[1070] 5821 TATATTATCA TATTATTTAT ATTATGTGCT TATTTCTGCT CTTATGGAAC CATCTCGATT
[1071] 5881 GTTAAATAT TTTTATATTT AATGGTATTA CAGGCGGTTA TTATATCCAT CATTAGTATT
[1072] 5941 TATATGACAA AACATATGG TATTGGTGAT TATTCAGCAC TAAGACATTA TTTTTTGGAG
[1073] 6001 AATGATTATG GTGATGTTTA TACATATGGA AGTGGTTTCT ATAGAGTTCA AATTAAAGGA
[1074] 6061 AATGCTCTCA TTCCATTTGC CTTTATGTTG CATATAGTCA TAAAAGATTA TTTCTATTAT
[1075] 6121 CGATTCAAAA ATACAATAAC CGTTATTCTG GCTATAGGTA CTATAGTGGC TGGTAATTTT
[1076] 6181 GCATATTTTG TTTGATATG CTTGTTTTTT ATGTATATTA TACTATGTTT TAAATCTAAC
[1077] 6241 TCACGATACG CTAAATTAAG GAAAATTATT TTTGGGGTTT TTCTTACTGT GATTCTCCCT
[1078] 6301 TTTTTTATTA CATATTCAAT TGAGTTGATA ATCATGAAAT CAAATGGAGC TGATTCTTCT
[1079] 6361 TTAGGAGTTA GATGGGATCA GTTTACTGTA TTAATTAATG ATCTTACAGA GTCTGTATCA
[1080] 6421 AATTTTGTTA TAGGTTCTGG TTTGGGTAAT GTCATCAAAA TTCAAACCTC TATCCGTGAT
[1081] 6481 TATAGTGCAT ATATATATTA TGAATTGCAG TCAGTTTATT TTTTAAATCA ACTTGGCGTT
[1082] 6541 ATTTTATTTA CTTTGTTTTT ATTAATTAAT CTCCTTCTCA CGATTAAAAAT CATAAAATAC
[1083] 6601 AGTGAGTTGT GTGTGCTATA TTTTCTATAT GTTTCTTATG CAATTACTAA TCCTTATATT
[1084] 6661 TTAGACTCTA ACCATGTTGC TGTAATAATT GTATTAGTGA CATTAAGTAA TGTTCATAAA
[1085] 6721 AAGATGAAAG CTAAATGAAG GTTTTAAGGT GAAGATGGAC ACTGTATATG CCGTTTTGGT
[1086] 6781 TGCTTACAAC CCAGAACATA ATGATTTAAA AAATGCGGTT GAATTATTGT TGAGACAAGT
[1087] 6841 TACTAAAGTT GTCGTTTGCA ATAACCTCTAC AAATGGTTAT AAATATGCTG AAAATTCTTC
[1088] 6901 AGGCGATGTA AAAATATTCA ATTTCAATGA TAATTTAGGC ATAGCAGAAG CCCAAAGTAT

[1089] 6961 AGGAATGAAA TGGGCTTTTG AAAATGGCGC TGATTTTATA TTGCAAATGG ATCAGGATAG
[1090] 7021 TATTCCTGAT CCTAAGATGG TAGAGCAGTT ACTTACTTGT TACAAAAAAT TGCTTAAACA
[1091] 7081 AAATGTCAAT GTTGGTTTAG TTGGTTCACA AGATTTTGAT AAAGTAACTG GTGAATTAAA
[1092] 7141 TAAAGCAAGG GTAAAAAAG GGAACCCT TACAGAAGTT TATTATGAGG TAGATAGTAC
[1093] 7201 ATTAAGTTCT GGCAGTCTAA TACCAAAAAA TAGTTGGTTG ATTGTTGGAG GAATGAAAGA
[1094] 7261 TGAGCTTTTT ATCGATGCGG TAGACCATGA ATATTGTTGG AGATTAAGAG CTGCTGGGTT
[1095] 7321 TAAAGTAATT AGGAATAAAA ATGCGTTACT TGCACATAGA CTTGGAGATG GGCGATTTAA
[1096] 7381 GATCTTAAAT ATTCTTTCTG TCGGTTTGCC AAGCCCATT TCGTCATTATT ATGCTACTCG
[1097] 7441 AAATATCTTT CTTTTATTAA ATAAAAATTA TGTACCCATC TACTGGAAAA TTTCTAGTCT
[1098] 7501 GGTAAATTA ATTGGAAAGG TTTTTTTATA TCCTATTTTC CTTCCAAATG GTAATAAAAG
[1099] 7561 GTTATATTTT TTTTAAAAAG GCATTAATGA CGGTTTAATG GGTGGAAGTG GTAAAAAGAA
[1100] 7621 ATGAATCATA GATTAGAAAA ATTCTCAGTT TTAATTAGCA TTTATAAAAA TGATCTACCG
[1101] 7681 CAATTTTTTG AGGTGGCTCT ACGCTCTATT TTTCACGATC AAACACTTAA GCCAGATCAA
[1102] 7741 ATAGTAATTG TTGCAGATGG AGAACTCCAT CAAACACACA TCGATATTAT AAATTCATTC
[1103] 7801 ATTGATGATG TTGGCAATAA AATAGTAACA TTTGTACCTT TACCTAGAAA TGTTGGATTG
[1104] 7861 GCTAATGCCT TAAATGAAGG ATTAAGGCT TGTAGGAATG AGTTAGTGGC AAGAATGGAT
[1105] 7921 GCTGATGATA TTTCTTTGCC TCATCGGTTT GAGAAACAAA TTTCTTTTAT GATTAATAAT
[1106] 7981 TCAGAAATAG ATGTATGTGG CAGTTTTATT GATGAAATTG AAAGTGTAC TGAGGAGTTT
[1107] 8041 ATTTCAACAC GCAAAGTGCC TCTCGAACAT AGAGAAATAG TAAATTCGC GAGGAAACGA
[1108] 8101 AGCGCAGTTA GCCATCCTTC TGTAATTTTT AGAAAGAATA CAGTATTAGC TGTTGGTGGT
[1109] 8161 TATCCTCCAT TCAGAAAATC TCAAGATTTT GCATTGTGGA GCCTATTAAT TGTACATAAT
[1110] 8221 GCAAGATTG CAAATCTTCC AGATATTTTA TAAAAATGC GAACTGGTCG TAATCTTATG
[1111] 8281 GCTCGACGTG GATTGTCATA TTTATTGTAC GAGTATAAAG TATTGTATTA TCAATATAAA
[1112] 8341 ATTGGTTTTA TTCGAAAAAA TGAATTAATA AGTAATGCTA TGTTGAGAAC ATTTTTTCGT
[1113] 8401 ATAATGCCAT CTAAATTAAA GGAGCTGATG TATTCAATCG TTAGGAATCG ATAATAATAA
[1114] 8461 TTTTCTGATT AAGTGTATG GATTTATTTT TATTAGGCAT ATTCTATAAT TAAGCATAAC
[1115] 8521 CCGCATACCA CCCAGCGGTA TCCTGACAGG AGTAAACAAT GTCAAAGCAA CAGATCGGCG
[1116] 8581 TCGTCGGTAT GGCAGTGATG GGGCGCAACC TTGCGCTCAA TATCGAAAGC CGTGGTTATA
[1117] 8641 CCGTCTCTAT TTTCAACCGT TCCCGTGAAA AGACCGAAGA AGTGATTACC GAAAATCCAG
[1118] 8701 GCAAGAACT GGTTCCTTAC TATACGGTGA AAGAATTTGT TGAATCTCTG GAAACGCCTC
[1119] 8761 GTCGCATCCT GTTAATGGTG AAAGCAGGTG CTGGCACGGA TGCTGCTATT GATTCCCTCA
[1120] 8821 AGCCATACCT CGATAAAGGT GACATCATCA TTGATGGTGG TAACACCTTC TTCCATGACA
[1121] 8881 CCATTCGTCG TAACCGTGAG CTTTCTGCAG AAGGCTTTAA CTTTATCGGT ACCGGTGTTT
[1122] 8941 CCGGTGGTGA AGAAGGTGCG CTGAAAGGTC CTTCCATTAT GCCTGGTGGG CAGAAAGAAG
[1123] 9001 CTTATGAACT GATTGCGCCG ATCCTGACCA AAATCGCCGC TGTGGCTGAA GACGGCGAAC
[1124] 9061 CGTGCGTTAC CTATATTGGT GCCGATGGTG CAGGTCATTA TGTGAAGATG GTTCAACACG
[1125] 9121 GTATTGAATA CGGTGATATG CAGCTGATTG CTGAAGCCTA TTCTCTGCTT AAAGGTGGCT
[1126] 9181 TGAACCTCAC CAACGAAGAA CTGGCGCAGA CCTTTACCGA GTGGAATAAC GGTGAACTGA
[1127] 9241 GCAGCTACCT GATCGACATC ACCAAAGATA TCTTCACCAA AAAAGATGAA GAGGGTAACT

[1128] 9301 ACCTGGTTGA TGTGATTCTG GATGAAGCAG CAAACAAAGG TACGGGCAAA TGGACCAGCC
 [1129] 9361 AGAGCGCGCT GGATCTCGGC GAACCGCTGT CGCTGATTAC CGAGTCTGTG TTTGCACGTT
 [1130] 9421 ATATCTCTTC TCTGAAAGAG CAGCGTGTG CCGCATCTAA AGTTCTCTCT GGCCCGCAAG
 [1131] 9481 CGCAGCCAGC TGGCGACAAT GCTGAGTTCA TCGAAAAAGT TCGCCGTGCG CTGTATCTGG
 [1132] 9541 GCAAAATCGT TTCTTACGCT CAGGGCTTCT CTCAGCTACG CGCTGCGTCT GAAGAGTACA
 [1133] 9601 ACTGGGATCT GAACTACGGT GAAATCGCGA AGATTTTCCG TGCTGGCTGC ATCATCCGTG
 [1134] 9661 CGCAGTTCCT GCAGAAAATC ACCGATGCTT ATGCCGAAAA TCCGCAGATC GCTAACCTGT
 [1135] 9721 TGCTGGCTCC TTA CTTC AAG CAAATTGCCG ATGACTACCA GCAGGCGCTG CGCGATGTCG
 [1136] 9781 TCGCTTACGC AGTACAGAAC GGTATCCCGG TGCCGACCTT CGCCGCTGCG GTTGCCTATT
 [1137] 9841 ACGACAGCTA CCGCGCCGCT GTTCTGCCTG CGAACCTGAT CCAGGCACAG CGTGACTATT
 [1138] 9901 TCGGTGCGCA TACTTATAAG CGCATTGATA AAGAAGGTGT GTTCCATACC GAATGGCTGG
 [1139] 9961 ATTAATCTGA TTAAATCAA TTAATCAAAG CAAGGCCCGG AGAAACCCTC CGGGCTTTTT
 [1140] 10021 TATTATACAA AGCGGCAGGT TAGGGCCTTT TTTTATAATT TATAGTTAAA AACGCGATAT
 [1141] 10081 AATACAGCGC CGCACAGCAG GATCGCTGCC TTGACAGTTC ATCTACATCA GCGTTAAAAA
 [1142] 10141 TCCCGCAGTA GATGAAGCTG TGGTGGTGGA TTAATGACCA CTCTAAATGT TTAACCGGAA
 [1143] 10201 GAAGTCAGAG CTAATGAAAA TAACAATTTT AGGAACAGGT TATGTTGGTC TTTCAAATGG
 [1144] 10261 TATTCTGATT GCGCAAAACC ACGAAGTGGT TGCCTGGAT ATCGTTCAGG CCAAAGTGGG
 [1145] 10321 CATGCTTAAC AAGAGGCAGT CACCGCTTGT TGATAAGGAG ATTGAAGAGT ATCTGGCGAC
 [1146] 10381 TAAAGATCTC AATTTCCGCG CTACGACAGA TAAGTATGAC GCGTATAAAA ATGCCGATTA
 [1147] 10441 CGTTATTATT GCCACACCTA CCGATTATGA TCCGAAAACA AATTATTTTA ATACCTCAAG
 [1148] 10501 CGTGGAAGCG GTCATTGCTG ATGTGACAGA AATTAATCCC AACGCGGTAA TGATTATAAA
 [1149] 10561 ATCAACTATC CCTGTTGGTT TTACAGAGTC CATTAAAGAA CGTTTTGGTA TTGAAAATGT
 [1150] 10621 GATCTTTTCG CCTGAGTTTT TGCGTGAAGG TAAAGCACTT TATGATAACT TACACCCATC
 [1151] 10681 ACGCATTGTG ATTGGCGAGC AGTCTGAACG CGCTAAACGT TTTGCTGCGT TATTACAGGA
 [1152] 10741 AGGCGCCATT AAGCAAGACA TACCAACATT GTTTACTGAC TCAACCGAGG CTGAGGCGAT
 [1153] 10801 TAACTTTTTT GCGAACACTT ATCTGGCGAT GCGTGTAGCG TATTTCAATG AACTTGATAG
 [1154] 10861 TTATGCTGAA AGCCTGGGAC TTAATTCACG CCAGATTATT GAGGGCGTAT GCCTTGACCC
 [1155] 10921 GCGTATCGGT AATCACTACA ACAACCCGTC ATTCGGTTAT GGTGGTTATT GTCTGCCGAA
 [1156] 10981 AGATACTAAG CAGTTACTGG CAAATTACCA GTCTGTGCCG AATAACCTGA TCTCGGCAAT
 [1157] 11041 TGTTGACGCC AACCGCACGC GCAAAGATTT TATTGCCGAT TCTATCCTTG CACGTAAACC
 [1158] 11101 GAAAGTTGTT GGCGTCTATC GTTTGATTAT GAAGAATGGT TCAGACAATT TTCGTGCTTC
 [1159] 11161 CTCGATTCAG GGTATTATGA AGCGAATCAA GGCGAAAGGT GTGCCTGTAA TCGTTTATGA
 [1160] 11221 GCCAGCTATG AAAGAGGACG ATTTTTTCCG GTCGCGCGTG GTACGTGATC TGGATGCGTT
 [1161] 11281 CAAACAAGAA GCTGATGTTA TTATTTCTAA CCGTATGTCT GCCGATCTGG CTGATGTAGC
 [1162] 11341 AGATAAAGTT TATACGCGCG ACTTGTTTGG CAATGATTAA TTATTTTGTT TCATTCTAAG
 [1163] 11401 AAAAGGCCCT AATAAATTAG GGCCTTTTCT TATGGTTTTG TAAATCAAAA CTTTATAGAA
 [1164] 11461 GTTACGATAC CATTCTACAA AGTTCTTTAC CCCTTCTTTA ACTGACGTTT CAGGTTTGAA
 [1165] 11521 TCCTATTACG TCATACAGTG CTTTTGTATC AGCACTGGTT TCCAGTACAT CACCGGGTTG
 [1166] 11581 GAGAGGCATC ATATTTTTGT TGGCTTCAAT ACCCAGAGCC TCTTCTAACG CATTGATATA

[1167] 11641 GTCCATCAAC TCCACAGGCG AACTATTACC AATGTTATAG ACACGATATG GTGCTGAACT
[1168] 11701 TGTTGCAGGC GAGCCTGTTT CTACAGCCCA CTGTGGGTTT TTTTCTGGAA TAACATCCTG
[1169] 11761 TAAGCGAATA ATAGCTTCGG CAATATCATC AATGTAAGTA AAGTCACGCT TCATTTTGCC
[1170] 11821 GAAGTTGTAA ACATCAATGC TTTTACCTTC CAGCATGGCT TTAGTGAATT TAAATAATGC
[1171] 11881 CATATCCGGA CGTCCCCATG GACCATAAAC CGTAAAGAAA CGCAGCCCTG TGGTCGGTAA
[1172] 11941 GCCATACAAA TGAGAATATG TATGGGCCAT GAGTTCATTC GCTTTTTTATG TTGCTGCATA
[1173] 12001 AAGCGAAACA GGATGATCTA CAGAGTCATC TGTAGAGAAA GGCATCTTGC GGTTCATGCC
[1174] 12061 ATAAACAGAA CTGGAGGAAG CGTAAAGTAG ATGCTGAACA TTATTATGGC GACATCCTTC
[1175] 12121 TAGTATGTTT AGGAATCCAA TCAGGTTTGC ATCTGCATAT GCATTGGGAT TTTCAAGAGA
[1176] 12181 GTAACGTACA CCGGCTTGCG CAGCGAGGTT TATTACGCGT TCGAACCGCT CGTCTGCAAA
[1177] 12241 CAGTGCCGCC ATTTTCTCAC GATCGGCCAG GTCAATTTTA TAAAACTGA AGTTGTCTGT
[1178] 12301 CTTGAGTAAA TCAAGTCGTG CTTGTTTGAG GTTGACATCG TAATAATCAT TTAAGTTGTC
[1179] 12361 AATGCCTACA ACCTGATGAC CAGCTGCAAG AAGCCGTTTA CTTAGATAGA AACCGATAAA
[1180] 12421 GCCAGCAGCT CCCGTAACCA GAAATTTTAT TTATAATCCT CGCTCAGGCT AGAATATAGC
[1181] 12481 CAATCTTCAT CTGGCATAAC TGAAAGTTAA ATTATACCGT TAGACAAGAA AAAAAGATAA
[1182] 12541 TCGGTATCAG TTCTAAACTT GGCTGTTTTT TCTGGTAACG TGCTCATTTT ACAATCAAAG
[1183] 12601 CTGTTCTAAG CTGACTATAC AAGCCGACGT CATTATCTCC AACCGTATGG CAGAAGAGCT
[1184] 12661 TAAGGATGTG GCAGACAAAG TCTACACCCG CGATCTCTTT GGCAGTGACT AACATCCTGT
[1185] 12721 TATCATGGCG ATTTTCGCCC TGATTCTCTT ATGTTCCCTT TGTAATAATT CATTATTTTT
[1186] 12781 ATCATTTATC CTATAGCATT CATGGCGATT ATCGCTAAAC TATGGCGGCG CGCCACGTGG
[1187] 12841 GATCCCCGGG TACCGAGCTC GAATTCGCCC TATAGTGAGT CGTATTACAA TTCACTGGCC
[1188] 12901 GTCGTTTTAC AACGTCGTGA CTGGGAAAAC CCTGGCGTTA CCCAACTTAA TCGCCTTGCA
[1189] 12961 GCACATCCCC CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG CCCGCACCGA TCGCCCTTCC
[1190] 13021 CAACAGTTGC GCAGCTGAAT GGCGAATGGC GCCTGATGCG GTATTTTCTC CTTACGCATC
[1191] 13081 TGTGCGGTAT TTCACACCGC ATATGGTGCA CTCTCAGTAC AATCTGCTCT GATGCCGCAT
[1192] 13141 AGTTAAGCCA GCCCCGACAC CCGCCAACAC CCGCTGACGC GAACCCCTTG CGGCCGCATC
[1193] 13201 GAATATAACT TCGTATAATG TATGCTATAC GAAGTTATTA GCGATGAGCT CGGACTTCCA
[1194] 13261 TTGTTTCATTC CACGGACAAA AACAGAGAAA GGAAACGACA GAGGCCAAAA AGCTCGCTTT
[1195] 13321 CAGCACCTGT CGTTTCCTTT CTTTTCAGAG GGTATTTTAA ATAAAAACAT TAAGTTATGA
[1196] 13381 CGAAGAAGAA CGGAAACGCC TTAAACCGGA AAATTTTCAT AAATAGCGAA AACCCGCGAG
[1197] 13441 GTCGCCGCCC CGTAACCTGT CGGATCACCG GAAAGGACCC GTAAAGTGAT AATGATTATC
[1198] 13501 ATCTACATAT CACAACGTGC GTGGAGGCCA TCAAACCACG TCAAATAATC AATTATGACG
[1199] 13561 CAGGTATCGT ATTAATTGAT CTGCATCAAC TTAACGTAAA AACAACTTCA GACAATACAA
[1200] 13621 ATCAGCGACA CTGAATACGG GGCAACCTCA TGTCCGAGCT CGCGAGCTCG TCGACAGCGA
[1201] 13681 CAACTTGCAG TCGGATGCAG CCCGGTTAAC GTGCCGGCAC GGCCTGGGTA ACCAGGTATT
[1202] 13741 TTGTCCACAT AACCGTGCGC AAAATGTTGT GGATAAGCAG GACACAGCAG CAATCCACAG
[1203] 13801 CAGGCATACA ACCGCACACC GAGGTTACTC CGTTCTACAG GTTACGACGA CATGTCAATA
[1204] 13861 CTTGCCCTTG ACAGGCATTG ATGGAATCGT AGTCTCACGC TGATAGTCTG ATCGACAATA
[1205] 13921 CAAGTGGGAC CGTGGTCCCA GACCGATAAT CAGACCGACA ACACGAGTGG GATCGTGGTC

[1206] 13981 CCAGACTAAT AATCAGACCG ACGATACGAG TGGGACCGTG GTCCCAGACT AATAATCAGA
[1207] 14041 CCGACGATAC GAGTGGGACC GTGGTTCCAG ACTAATAATC AGACCGACGA TACGAGTGGG
[1208] 14101 ACCGTGGTCC CAGACTAATA ATCAGACCGA CGATACGAGT GGGACCATGG TCCCAGACTA
[1209] 14161 ATAATCAGAC CGACGATACG AGTGGGACCG TGGTCCCAGT CTGATTATCA GACCGACGAT
[1210] 14221 ACGAGTGGGA CCGTGGTCCC AGACTAATAA TCAGACCGAC GATACGAGTG GGACCGTGGT
[1211] 14281 CCCAGACTAA TAATCAGACC GACGATACGA GTGGGACCGT GGTCCCAGTC TGATTATCAG
[1212] 14341 ACCGACGATA CAAGTGGAAC AGTGGGCCCA GAGAGAATAT TCAGGCCAGT TATGCTTTCT
[1213] 14401 GGCCTGTAAC AAAGGACATT AAGTAAAGAC AGATAAACGT AGACTAAAAAC GTGGTCGCAT
[1214] 14461 CAGGGTGCTG GCTTTTCAAG TTCCTTAAGA ATGGCCTCAA TTTTCTCTAT AACTCAGTT
[1215] 14521 GGAACACGAG ACCTGTCCAG GTTAAGCACC ATTTTATCGC CCTTATACAA TACTGTGCT
[1216] 14581 CCAGGAGCAA ACTGATGTCG TGAGCTTAAA CTAGTTCTTG ATGCAGATGA CGTTTTAAGC
[1217] 14641 ACAGAAGTTA AAAGAGTGAT AACTTCTTCA GCTTCAAATA TCACCCAGC TTTTTTCTGC
[1218] 14701 TCATGAAGGT TAGATGCCTG CTGCTTAAGT AATTCCTCTT TATCTGTAAA GGCTTTTTGA
[1219] 14761 AGTGCATCAC CTGACCGGGC AGATAGTTCA CCGGGGTGAG AAAAAAGAGC AACAACTGAT
[1220] 14821 TTAGGCAATT TGGCGGTGTT GATACAGCGG GTAATAATCT TACGTGAAAT ATTTTCCGCA
[1221] 14881 TCAGCCAGCG CAGAAATATT TCCAGCAAAT TCATTCTGCA ATCGGCTTGC ATAACGCTGA
[1222] 14941 CCACGTTTCAAT AAGCACTTGT TGGGCGATAA TCGTTACCCA ATCTGGATAA TGCAGCCATC
[1223] 15001 TGCTCATCAT CCAGCTCGCC AACCAGAACA CGATAATCAC TTTCGGTAAG TGCAGCAGCT
[1224] 15061 TTACGACGGC GACTCCCATC GGCAATTTCT ATGACACCAG ATACTCTTCG ACCGAACGCC
[1225] 15121 GGTGTCTGTT GACCAGTCAG TAGAAAAGAA GGGATGAGAT CATCCAGTGC GTCCTCAGTA
[1226] 15181 AGCAGCTCCT GGTACAGTTC ATTACCTGAC CATACCCGAG AGGTCTTCTC AACACTATCA
[1227] 15241 CCCCAGGACA CTTCAAGAGT AAACCTCACA TCCCGACCAC ATACAGGCAA AGTAATGGCA
[1228] 15301 TTACCGCGAG CCATTACTCC TACGCGCGCA ATTAACGAAT CCACCATCGG GGCAGCTGGT
[1229] 15361 GTCGATAACG AAGTATCTTC AACCGGTTGA GTATTGAGCG TATGTTTTGG AATAACAGGC
[1230] 15421 GCACGCTTCA TTATCTAATC TCCCAGCGTG GTTTAATCAG ACGATCGAAA ATTTTCATTGC
[1231] 15481 AGACAGGTTT CCAAAATAGAA AGAGCATTTT TCCAGGCACC AGTTGAAGAG CGTTGATCAA
[1232] 15541 TGGCCTGTTC AAAAAAGTT CTCATCCGGA TCTGACCTT ACCAACTTCA TCCGTTTCAC
[1233] 15601 GTACAACATT TTTTAGAACC ATGCTTCCCC AGGCATCCCC AATTTGCTCC TCCATCCACG
[1234] 15661 GGGACTGAGA GCCATTACTA TTGCTGTATT TGGTAAGCAA AATACGTACA TCAGGCTCGA
[1235] 15721 ACCCTTTAAG ATCAACGTTT TTGAGCAGAT CACGAAGCAT ATCGAAAAAC TGCAGTGCGG
[1236] 15781 AGGTGTAGTC AAACAACTCA GCAGGCGTGG GAACAATCAG CACATCAGCA GCACATACGA
[1237] 15841 CATTAATCGT GCCGATACCC AGGTTAGGCG CGCTGTCAAT AACTATGACA TCATAGTCAT
[1238] 15901 GAGCAACAGT TTCAATGGCC AGTCGGAGCA TCAGGTGTGG ATCGGTGGGC AGTTTACCTT
[1239] 15961 CATCAAATTT GCCCATTAAC TCAGTTTCAA TACGGTGCAG AGCCAGACAG GAAGGAATAA
[1240] 16021 TGTCAAGCCC CGGCCAGCAA GTGGGCTTTA TTGCATAAGT GACATCGTCC TTTTCCCCAA
[1241] 16081 GATAGAAAGG CAGGAGAGTG TCTTCTGCAT GAATATGAAG ATCTGGTACC CATCCGTGAT
[1242] 16141 ACATTGAGGC TGTTCCCTGG GGGTCGTTAC CTTCCACGAG CAAAAACAGT AGCCCTTCA
[1243] 16201 GAGCCAGATC CTGAGCAAGA TGAACAGAAA CTGAGGTTTT GTAAACGCCA CCTTTATGGG
[1244] 16261 CAGCAACCCC GATCACCGGT GGAAATACGT CTTACGACG TCGCAATCGC GTACCAAACA

[1245] 16321 CATCACGCAT ATGATTAATT TGTTC AATTG TATAACCAAC ACGTTGCTCA ACCCGTCTCTC
[1246] 16381 GAATTTCCAT ATCCGGGTGC GGTAGTCGCC CTGCTTCTCTG GGCATCTCTG ATAGCCTGAG
[1247] 16441 AAGAAACCCC AACTAAATCC GCTGCTTCAC CTATTCTCCA GCGCCGGGTT ATTTTCCTCG
[1248] 16501 CTTCCGGGCT GTCATCATT AACTGTGCAA TGGCGATAGC CTTGTCATT TCATGACCAG
[1249] 16561 CGTTTATGCA CTGGTTAAGT GTTTCCATGA GTTTCATTCT GAACATCCTT TAATCATTGC
[1250] 16621 TTTGCGTTTT TTTATTAAAT CTTGCAATTT ACTGCAAAGC AACAAACAAA TCGCAAAGTC
[1251] 16681 ATCAAAAAAC CGCAAAGTTG TTTAAAATAA GAGCAACACT ACAAAGGAG ATAAGAAGAG
[1252] 16741 CACATACCTC AGTCACTTAT TATCACTAGC GCTCGCCGCA GCCGTGTAAC CGAGCATAGC
[1253] 16801 GAGCGAACTG GCGAGGAAGC AAAGAAGAAC TGTTCGTCA GATAGCTCTT ACGCTCAGCG
[1254] 16861 CAAGAAGAAA TATCCACCGT GGGAAAAACT CCAGGTAGAG GTACACACGC GGATAGCCAA
[1255] 16921 TTCAGAGTAA TAAACTGTGA TAATCAACCC TCATCAATGA TGACGAACTA ACCCCCGATA
[1256] 16981 TCAGGTCACA TGACGAAGGG AAAGAGAAGG AAATCAACTG TGACAAACTG CCCTCAAATT
[1257] 17041 TGGCTTCCTT AAAAAATTACA GTTCAAAAAG TATGAGAAAA TCCATGCAGG CTGAAGGAAA
[1258] 17101 CAGCAAACT GTGACAAATT ACCCTCAGTA GGTCAGACA AATGTGACGA ACCACCCTCA
[1259] 17161 AATCTGTGAC AGATAACCT CAGACTATCC TGTCGTCATG GAAGTGATAT CGCGGAAGGA
[1260] 17221 AAATACGATA TGAGTCGTCT GCGGCCTTT CTTTTCTCA ATGTATGAGA GCGCATTGG
[1261] 17281 AGTTCTGCTG TTGATCTCAT TAACACAGAC CTGCAGGAAG CGGCGCGGA AGTCAGGCAT
[1262] 17341 ACGCTGGTAA CTTTGAGGCA GCTGGTAACG CTCTATGATC CAGTCGATTT TCAGAGAGAC
[1263] 17401 GATGCCTGAG CCATCCGGCT TACGATACTG ACACAGGGAT TCGTATAAAC GCATGGCATA
[1264] 17461 CGGATTGGTG ATTTCTTTTG TTTACTAAG CCGAACTGC GTAAACCGGT TCTGTAACCC
[1265] 17521 GATAAAGAAG GGAATGAGAT ATGGGTTGAT ATGTACACTG TAAAGCCCTC TGGATGGACT
[1266] 17581 GTGCGCACGT TTGATAAACC AAGGAAAAGA TTCATAGCCT TTTTCATCGC CGGCATCCTC
[1267] 17641 TTCAGGGCGA TAAAAACCA CTTCTTCCC CGCGAACTC TTCAATGCCT GCCGTATATC
[1268] 17701 CTTACTGGCT TCCGCAGAGG TCAATCCGAA TATTTCAGCA TATTTAGCAA CATGGATCTC
[1269] 17761 GCAGATACCG TCATGTTCTT GTAGGGTGCC ATCAGATTTT CTGATCTGGT CAACGAACAG
[1270] 17821 ATACAGCATA CGTTTTTGAT CCCGGGAGAG ACTATATGCC GCCTCAGTGA GGTGTTTTGA
[1271] 17881 CTGGACGATT CGCGGGCTAT TTTTACGTTT CTTGTGATTG ATAACCGCTG TTTCCGCCAT
[1272] 17941 GACAGATCCA TGTGAAGTGT GACAAGTTTT TAGATTGTCA CACTAAATAA AAAAGAGTCA
[1273] 18001 ATAAGCAGGG ATAACTTTGT GAAAAACAG CTTCTTCTGA GGGCAATTTG TCACAGGGTT
[1274] 18061 AAGGGCAATT TGTACAGAC AGGACTGTCA TTTGAGGGTG ATTTGTCACA CTGAAAGGGC
[1275] 18121 AATTTGTCAC AACACCTTCT CTAGAACCAG CATGGATAAA GGCCTACAAG GCGCTCTAAA
[1276] 18181 AAAGAAGATC TAAAACTAT AAAAAAATA ATTATAAAAA TATCCCGTG GATAAGTGGA
[1277] 18241 TAACCCCAAG GGAAGTTTTT TCAGGCATCG TGTGTAAGCA GAATATATAA GTGCTGTTCC
[1278] 18301 CTGGTGCTTC CTCGCTCACT CGACCGGGAG GGTTGAGAA GGGGGGCGAC CCCCTTCGG
[1279] 18361 CGTGCGCGGT CACGCGCACA GGGCGCAGCC CTGGTTAAAA ACAAGGTTTA TAAATATTGG
[1280] 18421 TTTAAAAGCA GGTAAAAGA CAGGTTAGCG GTGGCCGAAA AACGGGCGGA AACCTTGCA
[1281] 18481 AATGCTGGAT TTTCTGCCTG TGGACAGCCC CTCAAATGTC AATAGGTGCG CCCCTCATCT
[1282] 18541 GTCAGCACTC TGCCCTCAA GTGTCAAGGA TCGCGCCCT CATCTGTCAG TAGTCGCGCC
[1283] 18601 CCTCAAGTGT CAATACCGCA GGGCACTTAT CCCAGGCTT GTCCACATCA TCTGTGGGAA

[1284] 18661 ACTCGCGTAA AATCAGGCGT TTTCGCCGAT TTGCGAGGCT GGCCAGCTCC ACGTCGCCGG
 [1285] 18721 CCGAAATCGA GCCTGCCCCCT CATCTGTCAA CGCCGCGCCG GGTGAGTCGG CCCCTCAAGT
 [1286] 18781 GTCAACGTCC GCCCCTCATC TGTCAGTGAG GGCCAAGTTT TCCGCGAGGT ATCCACAACG
 [1287] 18841 CCGGCGGCCG GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 [1288] 18901 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCGAGGGCTT CGCCCTGTCTG
 [1289] 18961 CTCGACTGCG GCGAGCACTA CTGGCTGTAA AAGGACAGAC CACATCATGG TTCTGTGTTC
 [1290] 19021 ATTAGGTTGT TCTGTCCATT GCTGACATAA TCCGCTCCAC TTCAACGTAA CACCGCACGA
 [1291] 19081 AGATTTCTAT TGTTCCTGAA GGCATATTCA AATCGTTTTT GTTACCGCTT GCAGGCATCA
 [1292] 19141 TGACAGAACA CTACTIONCTA TAAACGCTAC ACAGGCTCCT GAGATTAATA ATGCGGATCT
 [1293] 19201 CTACGATAAT GGGAGATTTT CCCGACTGTT TCGTTCGCTT CTCAGTGGAT AACAGCCAGC
 [1294] 19261 TTCTCTGTTT AACAGACAAA AACAGCATAT CCACTCAGTT CCACATTTCC ATATAAAGGC
 [1295] 19321 CAAGGCATTT ATTCTCAGGA TAATTGTTTC AGCATCGCAA CCGCATCAGA CTCCGGCATC
 [1296] 19381 GCAAACGTCA CCCGGTGCCG GGCAGCCACA TCCAGCGCAA AAACCTTCGT GTAGACTTCC
 [1297] 19441 GTTGAACCTGA TGGACTTATG TCCCATCAGG CTTTGCAGAA CTTTCAGCGG TATACCGGCA
 [1298] 19501 TACAGCATGT GCATCGCATA GGAATGGCGG AACGTATGTG GTGTGACCGG AACAGAGAAC
 [1299] 19561 GTCACACCGT CAGCAGCAGC GGCGGCAACC GCCTCCCCAA TCCAGGTCCT GACCGTTCTG
 [1300] 19621 TCCGTCACCT CCCAGATCCG CGCTTTCTCT GTCCCTCCCTG TGCAGCGGTT ACGCCGCTCC
 [1301] 19681 ATGAGCTTAT CGCGAATAAA TACCTGTGAC GGAAGATCAC TTCGAGAAT AAATAAATCC
 [1302] 19741 TGGTGTCCCT GTTGATACCG GGAAGCCCTG GGCCAACTTT TGGCGAAAAT GAGACGTTGA
 [1303] 19801 TCGGCACGTA AGAGGTTCCT ACTTTCACCA TAATGAAATA AGATCACTAC CGGGCGTATT
 [1304] 19861 TTTTGAGTTA TCGAGATTTT CAGGAGCTAA GGAAGCTAAA ATGGAGAAAA AAATCACTGG
 [1305] 19921 ATATACCACC GTTGATATAT CCCAATGGCA TCGTAAAGAA CATTTTGAGG CATTTTCAGTC
 [1306] 19981 AGTTGCTCAA TGTACCTATA ACCAGACCGT TCAGCTGGAT ATTACGGCTT TTTTAAAGAC
 [1307] 20041 CGTAAAGAAA AATAAGCACA AGTTTTATCC GGCCTTTATT CACATTCTTG CCCGCTGAT
 [1308] 20101 GAATGCTCAT CCGGAATTTT GTATGGCAAT GAAAGACGGT GAGCTGGTGA TATGGGATAG
 [1309] 20161 TGTTACCCCT TGTTACACCG TTTTCCATGA GCAAACCTGAA ACGTTTTTCAT CGCTCTGGAG
 [1310] 20221 TGAATACCAC GACGATTTCC GGCAGTTTCT ACACATATAT TCGCAAGATG TGGCGTGTGA
 [1311] 20281 CGGTGAAAAC CTGGCCTATT TCCCTAAAGG GTTTATTGAG AATATGTTTT TCGTCTCAGC
 [1312] 20341 CAATCCCTGG GTGAGTTTCA CCAGTTTTGA TTTAAACGTG GCCAATATGG ACAACTTCTT
 [1313] 20401 CGCCCCCGTT TTCACCATGG GCAAATATTA TACGCAAGGC GACAAGGTGC TGATGCCGCT
 [1314] 20461 GGCGATTCAG GTTCATCATG CCGTTTGTGA TGGCTTCCAT GTCGGCAGAA TGCTTAATGA
 [1315] 20521 ATTACAACAG TACTGCGATG AGTGGCAGGG CGGGGCGTAA TTTTTTTAAG GCAGTTATTG
 [1316] 20581 GTGCCCTTAA ACGCCTGGTT GCTACGCCTG AATAAGTGAT AATAAGCGGA TGAATGGCAG
 [1317] 20641 AAATTCGATG ATAAGCTGTC AAACATGAGA ATTGGTCGAC GGCCCGGGCG GCCGCAAGGG
 [1318] 20701 GTTCGCGTTG GCCGATTCAT TAATGCAGCT GGCACGACAG GTTCCCGAC TGGAAGCGG
 [1319] 20761 GCAGTGAGCG CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC
 [1320] 20821 ACTTTATGCT TCCGGCTCGT ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTACACAG
 [1321] 20881 GAAACAGCTA TGACCATGAT TACGCCAAGC TATTTAGGTG AGACTATAGA ATACTCAAGC
 [1322] 20941 TTGCATGCCT GCAGGTCGAC TCTAGAGGAT CCCACGACGT CG

[1323] SEQ ID NO:29-pCC1F0S cut(pF0S)和弗氏志贺氏菌(*S.flexneri*)60-抗原(有Z3206)的核苷酸序列

[1324] 基因座 pF0S cut和O-抗原cut(Z3206+)

[1325] 定义用Z3206Nhe和wzzAscI扩增并用NheI和AscI切割的反向弗氏志贺氏菌(*S.flexneri*)60抗原簇连接到具有MCS盒并用NheI和AscI切割的pCC1F0S

特征	位置/合格者
CDS	补体(370..396) /标记物=wzz'
CDS	748..1752 /标记物=uge
CDS	补体(1818..3011) /标记物=ugd
[1326] CDS	补体(3233..4639) /标记物=gnd
CDS	补体(4744..5577) /标记物=wfbZ
CDS	补体(5574..6443) /标记物=wfbY
CDS	补体(6460..7647) /标记物=wzy
CDS	补体(7703..8935)

		/标记物=wzx
	CDS	补体(8932..9489)
		/标记物=rmlC
	CDS	补体(9494..10372)
		/标记物=rmlA
	CDS	补体(10430..11329)
		/标记物=rmlD
	CDS	补体(11329..12414)
		/标记物=rmlB
	CDS	补体(12787..13680)
		/标记物=galF
[1327]	CDS	补体(13912..14907)
		/标记物=Z3206
	CDS	补体(15065..15097)
		/标记物='wcaM
	CDS	补体(15525..16184)
		/标记物=cat
	CDS	16403..16750
		/标记物=redF
	CDS	18145..18900
		/标记物=repE
	CDS	19479..20654
		/标记物=parA
	CDS	20654..21625
		/标记物=parB
[1328]	长度:22887bp	
[1329]	类型:DNA 环状UNA	
[1330]	序列:	
[1331]	1	GCGGCCGCAA GGGGTTCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACATATG
[1332]	61	CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT
[1333]	121	GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGCTGCGCA ACTGTTGGGA
[1334]	181	AGGGCGATCG GTGCGGGCCT CTTCGCTATT ACGCCAGCTG GCGAAAGGGG GATGTGCTGC

[1335] 241 AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCAGTCA CGACGTTGTA AAACGACGGC
 [1336] 301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG CTCGGTACCC GGGGATCCCA
 [1337] 361 CGTGGCGCGC CGCCATAGTT TAGCGATAAT CGCCATGAAT GCTATAGGAT AAATGATAAA
 [1338] 421 AATAATGAAT TATTACAAAG GGAACATAAG AGAATCAGGG CGAAAAATCGC CATGATAACA
 [1339] 481 GGATGTTAGT CACTGCCAAA GAGATCGCGG GTGTAGACTT TGTCTGCCAC ATCCTTAAGC
 [1340] 541 TCTTCTGCCA TACGGTTGGA GATAATGACG TCGGCTTGTA TAGTCAGCTT AGAACAGCTT
 [1341] 601 TGATTGTAAA ATGAGCACGT TACCAGAAAA AACAGCCAAG TTTAGAACTG ATACCGATTA
 [1342] 661 TCTTTTTTTC TTGTCTAACG GTATAATTTA ACTTTCAGTT ATGCCAGATG AAGATTGGCT
 [1343] 721 ATATTCTAGC CTGAGCGAGG ATTATAAATG AAATTTCTGG TTACGGGAGC TGCTGGCTTT
 [1344] 781 ATCGGTTTCT ATCTAAGTAA ACGGCTTCTT GCAGCTGGTC ATCAGGTTGT AGGCATTGAC
 [1345] 841 AACTTAAATG ATTATTACGA TGTCAACCTC AAACAAGCAC GACTTGATTT ACTCAAGCAC
 [1346] 901 GACAACTTCA GTTTTTATAA AATTGACCTG GCCGATCGTG AGAAAAATGGC GGCAGTGT
 [1347] 961 GCAGACGAGC GGTTCGAACG CGTAATAAAC CTCGCTGCGC AAGCCGGTGT ACGTTACTCT
 [1348] 1021 CTTGAAAATC CCAATGCATA TGCAGATGCA AACCTGATTG GATTCCTGAA CATACTAGAA
 [1349] 1081 GGATGTCGCC ATAATAATGT TCAGCATCTA CTTTACGCTT CCTCCAGTTC TGTTTATGGC
 [1350] 1141 ATGAACCGCA AGATGCCTTT CTCTACAGAT GACTCTGTAG ATCATCCTGT TTCGCTTTAT
 [1351] 1201 GCAGCAACTA AAAAAAGCGA TGAACATCATG GCCCATACAT ATTCTCATTT GTATGGCTTA
 [1352] 1261 CCGACCACAG GGCTGCGTTT CTTTACGGTT TATGGTCCAT GGGGACGTCC GGATATGGCA
 [1353] 1321 TTATTTAAAT TCACTAAAGC CATGCTGGAA GGTAAGCA TTGATGTTTA CAACTTCGGC
 [1354] 1381 AAAATGAAGC GTGACTTTAC TTACATTGAT GATATTGCCG AAGCTATTAT TCGCTTACAG
 [1355] 1441 GATGTTATTC CAGAAAAAAA CCCACAGTGG GCTGTAGAAA CAGGCTCGCC TGCAACAAGT
 [1356] 1501 TCAGCACCAT ATCGTGTCTA TAACATTGGT AATAGTTCGC CTGTGGAGTT GATGGACTAT
 [1357] 1561 ATCAATGCGT TAGAAGAGGC TCTGGGTATT GAAGCCAACA AAAATATGAT GCCTCTCCAA
 [1358] 1621 CCCGGTGATG TACTGGAAAC CAGTGCTGAT ACAAAGCAC TGTATGACGT AATAGGATTC
 [1359] 1681 AAACCTGAAA CGTCAGTTAA AGAAGGGGTA AAGAACTTTG TAGAATGGTA TCGTAACTTC
 [1360] 1741 TATAAAGTTT GATTTTACAA AACCATAAGA AAAGGCCCTA ATTTATTAGG GCCTTTTCTT
 [1361] 1801 AGAATGAAAC AAAATAATTA ATCATTGCCA AACAAGTCGC GCGTATAAAC TTTATCTGCT
 [1362] 1861 ACATCAGCCA GATCGGCAGA CATACGTTA GAAATAATAA CATCAGCTTC TTGTTTGAAC
 [1363] 1921 GCATCCAGAT CACGTACCAC GCGCGACCGG AAAAAATCGT CCTCTTTCAT AGCTGGCTCA
 [1364] 1981 TAAACGATTA CAGGCACACC TTTCGCCTTG ATTCGCTTCA TAATACCCTG AATCGAGGAA
 [1365] 2041 GCACGAAAAT TGTCTGAACC ATTCTTCATA ATCAAACGAT AGACGCCAAC AACTTTCGGT
 [1366] 2101 TTACGTGCAA GGATAGAATC GGCAATAAAA TCTTTGCGCG TCGGTTGGC GTCAACAATT
 [1367] 2161 GCCGAGATCA GGTATTTCGG CACAGACTGG TAATTTGCCA GTAAGTCTT AGTATCTTTC
 [1368] 2221 GGCAGACAAT AACCACCATA ACCGAATGAC GGGTTGTTGT AGTGATTACC GATACGCGG
 [1369] 2281 TCAAGGCATA CGCCCTCAAT AATCTGGCGT GAATTAAGTC CCAGGCTTTC AGCATAACTA
 [1370] 2341 TCAAGTTCAT TGAAATACGC TACACGCATC GCCAGATAAG TGTTGCAAAA AAGTTTAATC
 [1371] 2401 GCCTCAGCCT CGGTTGAGTC AGTAAACAAT GTTGGTATGT CTTGCTTAAT GGCGCCTTCC
 [1372] 2461 TGTAATAACG CAGCAAAACG TTTAGCGCGT TCAGACTGCT CGCCAATCAC AATGCGTGAT
 [1373] 2521 GGGTGTAAGT TATCATAAAG TGCTTTACCT TCACGCAAAA ACTCAGGCGA AAAGATCACA

[1374] 2581 TTTTCAATAC CAAAACG TTC TTTAATGGAC TCTGTAAAAC CAACAGGGAT AGTTGATTTT
[1375] 2641 ATAATCATT CCGCGTTGGG ATTAATTTCT GTCACATCAC GAATGACCGC TTCCACGCTT
[1376] 2701 GAGGTATTAA AATAATTTGT TTTCGGATCA TAATCGGTAG GTGTGGCAAT AATAACGTAA
[1377] 2761 TCGGCATTTT TATACGCGTC ATACTTATCT GTCGTAGCGC GGAAATTGAG ATCTTTAGTC
[1378] 2821 GCCAGATACT CTTCAATCTC CTTATCAACA AGCGGTGACT GCCTCTTGTT AAGCATGTCC
[1379] 2881 ACTTTGGCCT GAACGATATC CAGTGCAACC ACTTCGTGGT TTTGCGCAAT CAGAATACCA
[1380] 2941 TTTGAAAGAC CAACATAACC TGTTCTGAA ATTGTTATTT TCATTAGCTC TGA CTCTCTC
[1381] 3001 CGGTAAACA TTTAGAGTGG TCATTAATCC ACCACCACAG CTTTCATCTAC TGCGGGATTT
[1382] 3061 TTAACGCTGA TG TAGATGAA CTGTCAAGGC AGCGATCCTG CTGTGCGGCG CTGTATTATA
[1383] 3121 TCGCGTTTTT AACTATAAAT TATAAAAAAA GGCCCTAACC TGCCGCTTTG TATAATAAAA
[1384] 3181 AAGCCCGGAG GGTTCCTCCG GGCCTTGCTT TGATTAATTG ATTTAAATCA GATTAATCCA
[1385] 3241 GCCATTCCGT ATGGAACACA CCTTCTTTAT CAATGCGCTT ATAAGTATGC GCACCGAAAT
[1386] 3301 AGTCACGCTG TGCCTGGATC AGGTTGCGAG GCAGAACAGC GCGCGGTAG CTGTGTAAT
[1387] 3361 AGGCAACCGC AGCGGCGAAG GTCGGCACC GGTACCGTT CTGTACTGCG TAAGCGACGA
[1388] 3421 CATCGCGCAG CGCCTGCTGG TAGTCATCGG CAATTTGCTT GAAGTAAGGA GCCAGCAACA
[1389] 3481 GGT TAGCGAT CTGCGGATTT TCGGCATAAG CATCGGTGAT TTTCTGCAGG AACTGCGCAC
[1390] 3541 GGATGATGCA GCCAGCACGG AAAATCTTCG CGATTTTACC GTAGTTTACA TCCCAGTTGT
[1391] 3601 ACTCTTACA CGCAGCGCGT AGCTGAGAGA AGCCCTGAGC GTAAGAAACG ATTTTGCCCA
[1392] 3661 GATACAGCGC ACGGCGAAT TTTTCGATGA ACTCAGCATT GTCGCCAGCT GGCTGCGCTT
[1393] 3721 GCGGGCCAGA GAGAACTTTA GATGCGGCAA CACGCTGCTC TTTACAGAGAA GAGATATAAC
[1394] 3781 GTGCAAACAC AGACTCGGTA ATCAGCGACA GCGGTTGCGC GAGATCCAGC GCGCTCTGGC
[1395] 3841 TGGTCCATTT GCCCGTACCT TTGTTTGCTG CTTTCATCCAG AATCACATCA ACCAGGTAGT
[1396] 3901 TACCCTCTTC ATCTTTTTTG GTGAAGATAT CTTTGGTGAT GTCGATCAGG TAGCTGCTCA
[1397] 3961 GTTCACCGTT ATTCCACTCG GTAAAGGTCT GCGCCAGTTC TTCGTTGGTG AGGTTCAAGC
[1398] 4021 CACCTTTAAG CAGAGAATAG GCTTCAGCAA TCAGCTGCAT ATCACCCTAT TCAATACCGT
[1399] 4081 TGTGAACCAT CTTACATAA TGACCTGCAC CATCGGCACC AATATAGGTA ACGCACGGTT
[1400] 4141 CGCCGTCTTC AGCCACAGCG GCGATTTTG TCCAGGATCG CGCAATCAGT TCATAAGCTT
[1401] 4201 CTTTCTGCCC ACCAGGCATA ATGGAAGGAC CTTTCAGCGC ACCTTCTTCA CCACCGGAAA
[1402] 4261 CACCGGTACC GATAAAGTTA AAGCCTTCTG CAGAAAGCTC ACGGTTACGA CGAATGGTGT
[1403] 4321 CATGGAAGAA GGTGTTACCA CCATCAATGA TGATGTCACC TTTATCGAGG TATGGCTTGA
[1404] 4381 GGAATCAAT AGCAGCATCC GTGCCAGCAC CTGCTTTCAC CATTAACAGG ATGCGACGAG
[1405] 4441 GCGTTTCCAG AGATTCAACA AATTCTTTCA CCGTATAGTA AGGAACCACT TTCTTGCTG
[1406] 4501 GATTTTCCGT AATCACTTCT TCGGTCTTTT CACGGGAACG GTTGAAAATA GAGACGGTAT
[1407] 4561 AACCACGGCT TTCGATATTG AGCGCAAGGT TGCGCCCAT CACTGCCATA CCGACGACGC
[1408] 4621 CGATCTGTTG CTTTGACATT GTTTACTCCT GTCAGGATAC CGCTGGGTGG TATGCGGGTT
[1409] 4681 ATGCTTAATT ATAGAATATG CCTAATAAAA ATAAATCCAT AACACTTAAT CAGAAAATTA
[1410] 4741 TTATTATCGA TTCCTAACGA TTGAATACAT CAGCTCCTTT AATTTAGATG GCATTATACG
[1411] 4801 AAAAAATGTT CTCAACATAG CATTACTTAT TAATTCATTT TTTGCAATAA AACCAATTTT
[1412] 4861 ATATTGATAA TACAATACTT TATACTCGTA CAATAAATAT GACAATCCAC GTCGAGCCAT

[1413] 4921 AAGATTACGA CCAGTTCGCA TTTTAAATAA AATATCTGGA AGATTTGCAA ATCTTGCAAT
 [1414] 4981 ATGTACAATT AATAGGCTCC ACAATGCAAA ATCTTGAGAT TTTCTGAATG GAGGATAACC
 [1415] 5041 ACCAACAGCT AATACTGTAT TCTTTCTAAA AATTACAGAA GGATGGCTAA CTGCGCTTCG
 [1416] 5101 TTTCCCTCGCG AATTTAACTA TTTCTCTATG TTCGAGAGGC ACTTTGCGTG TTGAAATAAA
 [1417] 5161 CTCCTCAGTA ACAGTTTCAA TTTTCATCAAT AAAACTGCCA CATACATCTA TTTCTGAATT
 [1418] 5221 ATTAATCATA AAAGAAATTT GTTTCTCAAA CCGATGAGGC AAAGAAATAT CATCAGCATC
 [1419] 5281 CATTCTTGCC ACTAACTCAT TCCTACAAGC CTTTAATCCT TCATTTAAGG CATTAGCCAA
 [1420] 5341 TCCAACATTT CTAGGTAAAG GTACAAATGT TACTATTTTA TTGCCAACAT CATCAATGAA
 [1421] 5401 TGAATTTATA ATATCGATGT GTGTTTGATG GAGTTCTCCA TCTGCAACAA TTACTATTTG
 [1422] 5461 ATCTGGCTTA AGTGTTTGAT CGTGAAAAAT AGAGCGTAGA GCCACCTCAA AAAATTGCGG
 [1423] 5521 TAGATCATTT TTATAAATGC TAATTA AAAAC TGAGAATTTT TCTAATCTAT GATTCATTTT
 [1424] 5581 ATTTTACCAC TTCGACCCAT TAAACCGTCA TTAATGCCTT TTA AAAAAA ATATAACCTT
 [1425] 5641 TTATTACCAT TTGGAAGGAA AATAGGATAT AAAAAACCT TTCCAATTAA TTTAACCAGA
 [1426] 5701 CTAGAAATTT TCCAGTAGAT GGGTACATAA TTTTATTTA ATAAAAGAAA GATATTTTGA
 [1427] 5761 GTAGCATAAT AATGACGAAA TGGGCTTGGC AAACCGACAG AAAGAATATT TAAGATCTTA
 [1428] 5821 AATCGCCCAT CTCCAAGTCT ATGTGCAAGT AACGCATTTT TATTCCTAAT TACTTTAAAC
 [1429] 5881 CCAGCAGCTC TTAATCTCCA ACAATATTCA TGGTCTACCG CATCGATAAA AAGCTCATCT
 [1430] 5941 TTCATTCCCTC CAACAATCAA CCAACTATTT TTTGGTATTA GACTGCCAGA ACTTAATGTA
 [1431] 6001 CTATCTACCT CATAATAAAC TTCTGTAAGT GGTTTCCCTT TTTTACCCT TGCTTTATTT
 [1432] 6061 AATTCACCAG TTACTTTATC AAAATCTTGT GAACCAACTA AACCAACATT GACATTTTGT
 [1433] 6121 TTAAGCAATT TTTTGTAACA AGTAAGTAAC TGCTCTACCA TCTTAGGATC AGGAATACTA
 [1434] 6181 TCCTGATCCA TTTGCAATAT AAAATCAGCG CCATTTTCAA AAGCCCATTT CATTCCTATA
 [1435] 6241 CTTTGGGCTT CTGCTATGCC TAAATTATCA TTGAAATTGA ATATTTTAC ATCGCCTGAA
 [1436] 6301 GAATTTTCAG CATATTTATA ACCATTTGTA GAGTTATTGC AAACGACAAC TTTAGTAACT
 [1437] 6361 TGTCTCAACA ATAATTCAAC CGCATTTTTT AAATCATTAT GTTCTGGGT GTAAGCAACC
 [1438] 6421 AAAACGGCAT ATACAGTGTC CATCTTCACC TTA AACCTT CATTTAGCTT TCATCTTTTT
 [1439] 6481 TAGAACATTA CTTAATGTCA CTAATACAAT TATTACAGCA ACATGGTTAG AGTCTAAAAT
 [1440] 6541 ATAAGGATTA GTAATTGCAT AAGAAACATA TAGAAAATAT AGCACACACA ACTCACTGTA
 [1441] 6601 TTTTATGATT TTAATCGTGA GAAGGAGATT AATTAATAAA AACAAAGTAA ATAAAATAAC
 [1442] 6661 GCCAAGTTGA TTTAAAAAAT AAAGTACTG CAATTCATAA TATATATATG CACTATAATC
 [1443] 6721 ACGGATAGGA GTTTGAATTT TGATGACATT ACCCAAACCA GAACCTATAA CAAAATTTGA
 [1444] 6781 TACAGACTCT GTAAGATCAT TAATTAATAC AGTAAACTGA TCCCATCTAA CTCCTAAAGA
 [1445] 6841 AGAATCAGCT CCATTTGATT TCATGATTAT CAACTCAATT GAATATGTAA TAAAAAAGG
 [1446] 6901 GAGAATCACA GTAAGAAAAA CCCCCAAAAT AATTTTCCTT AATTTAGCGT ATCGTGAGTT
 [1447] 6961 AGATTTAGAA CATAGTATAA TATACATAAA AAACAAGCAT ATCGAAACAA AATATGCAAA
 [1448] 7021 ATTACCAGCC ACTATAGTAC CTATAGCCAG AATAACGGTT ATTGTATTTT TGAATCGATA
 [1449] 7081 ATAGAAATAA TCTTTTATGA CTATATGCAA CATAAAGGCA AATGGAATGA GAGCATTTCC
 [1450] 7141 TTTAATTTGA ACTCTATAGA AACCATTCC ATATGTATAA ACATCACCAT AATCATTCTC
 [1451] 7201 CAAAAAATAA TGTCTTAGTG CTGAATAATC ACCAATACCA TATGTTTTTG TCATATAAAT

[1452] 7261 ACTAATGATG GATATAATAA CCGCCTGTAA TACCATTAAA TATAAAAATA TTTTAACAAT
[1453] 7321 CGAGATGGTT CCATAAGAGC AGAAATAAGC ACATAATATA AATAATATGA TAATATAAAA
[1454] 7381 CCTAATTATT ATCGCTATAT CGTTACCCTT GATATAGGAA TAAATAAAAT TTATAAAAAG
[1455] 7441 AGCTAATAGA AATATTAAAA TAACAGGATA GTGATATATT CCGTTTGCAA TTTTCTTTGT
[1456] 7501 AAATGACATG ATACAAAGAC ATAAAAACCC CTCCATAATC CAACTATATT GAATAAATGG
[1457] 7561 AAAGCTACGT GTAAGGAAAA ATATAAACCC AAAAAACAAA AGAACACTTA AACTTTTGTG
[1458] 7621 TTTTGAGTTA TAAAAATCAG AAGTCATGTT TGCACCTCTAA TTAGATGGGC TTGAGGAAGT
[1459] 7681 AATCCCTAAA ATCAATTCGC TATTAATATT TCGTATCAAT TAATAATAAT ATCAAAAAAT
[1460] 7741 CTAACGATGT TCTTACAGAC CATGCTATTG CGGCTCCAAC AATTCCCCAA TGATAAATAA
[1461] 7801 AAATATATAA TATGCATAAA TATGGGATAA CTTCGAGCAA ATGAATAATA GCTGTAATTT
[1462] 7861 TTGATCTTCC ACTAGCCTGA ACTGAAACAA ATGGGATTG TGCAATGCAA TTAaaaaaga
[1463] 7921 AACCTATTGC AAGAATTTTT AATACTATAC CTGGCGTCCC ATGATATGTA GGTCCCATCC
[1464] 7981 AAGCGGACAT TATAAAATCT GATAAAATAA TTATCAACAT TACAATTGGA AGTATACCAA
[1465] 8041 TAACCATTAT AAAATATGAT AATATTTTAG TTTGCTTTAC CGATTGCAAT TCTGAACTTA
[1466] 8101 ATCTTGGAAG AATAGCTCTG GACAACGCAC TTGGTAATAT CGTTAAGCGT TGTATACCTT
[1467] 8161 CAGACGGAGC AGTATAAAAA GAAACTTTAT CAGCCCCCAC AATGTGTGAA AGAATAAAAC
[1468] 8221 GATCCATATA TGTCAATAA GGGCTAATAA TATTGCTAAC TGTTATCCAG CTTCCAAAGC
[1469] 8281 CGATTAATCT TTAACTGTT ACAATTTTTA CAGACAGCCC AGATGATATT ATTAGTTTTC
[1470] 8341 GACTAAATAT AAAGGTCAC TAAAGTGATA AGACTCTTGC CATAACTAAA CCATATATAG
[1471] 8401 CACTTAGTAA TCCTCCATGA AAAAAACAGA AAATCACTGG TAATCCAGCC ACAAAGAGT
[1472] 8461 TGTTAATTGA TTTTATTAAA TTTACTTTTC TGAACCTTTC CATCCCCTCA AAAATCCCCA
[1473] 8521 ACCAGACTTG GTTAAACAAG TATAAGGGTA TGGTAGCTGA AATAATATAT ATTGCTTTGA
[1474] 8581 CAGATTCTAC AACATGATTC GCGTTAATGT TTAATAATTT AACAAATTACA TTGCTACTCA
[1475] 8641 AAAATAGTAC ACTACCGCCA ATCAAGCCCA ATATAGTTAG AATTACCGTT GAAGTTGAAA
[1476] 8701 TGATCGCTCT TAATTCCTTA TGAACATTTT TATATATTGA TACTTCTCTT ATAACAGCTC
[1477] 8761 TGGTCAATCC AGCATCAAAA ATACTTGCAT ATCCAATAA GGCAATAGCT AACGTAAAAA
[1478] 8821 GGCCAAATTG CTCGGTCCCT AGAATTCTAG ACAGTATACC TAACGCAGGA ATTGCTATTA
[1479] 8881 ATGATGGTAT AATATACCCA CTTATATTCC ATAAAGTATT CTTTACAATA CTCACAAAAA
[1480] 8941 TAATTCCTTC ATGTTATGCA ATTCTTTAGC CCTTGCATCT TTAATCGATA AAATATAATT
[1481] 9001 ATTATGTTCT ATCGTCGGCC ATTTTATGCT CAGAATAGGA TCATTCCATA CAATCCCTCT
[1482] 9061 ATCACTATCA GGATGATAAT AGTTCGTCTG TTTATATAAA AATTCGCAG TCTCGCTCAG
[1483] 9121 CACCAAAAAA CCATGTGCAA ATCCCTCAGG GATCCACAAT TGCCGCTTAT TCTCAGCAGA
[1484] 9181 TAAATTCACC CCAACCCATT TACCAAAGGT AGGCGACGAT TTACGAATAT CAACAGCTAC
[1485] 9241 ATCAAAAACC TCACCAACAA CGCAACGTAC CAGTTTCCCT TGCGCATAAG GTTCTAACTG
[1486] 9301 ATAATGCAGC CCGCGTAAAA CACCTTTACT AGACTTCGAA TGGTTATCCT GAACAAATTC
[1487] 9361 AACCTTACGT CCTACAGCTT CTTGAAAAAC TTTCTGATTA AAGCTTTCCA TAAAGAAACC
[1488] 9421 ACGCTCATCA CCAAAAACTT TCGGCTCGAA AATTAACACA TCAGGAATTT CTGTTTTAAT
[1489] 9481 TACGTTTATT TTATTAATAA CCTTTAATCA TTTTCAGCAG ATACTGTCCA TAAGCATTTT
[1490] 9541 TTTTCAGCGG CTCCGCTAAT GCTTTCACCT GTTCAGCATC AATAAACCCCT TTACGGTAAG

[1491] 9601 CAATTTCTTC TGGGCAGGAA ACCTTTAGTC CCTGGCGCTC TTCAATGGTG GCAATGAAGT
[1492] 9661 TGCTTGCTTC AATAAGACTC TGATGTGTCC CCGTATCCAG CCATGCATAA CCACGCCCCA
[1493] 9721 TCATGGCAAC GGATAAACGC CCCTGTTCCA TATAAATACG GTTAATATCG GTAATTTCCA
[1494] 9781 GTTCACCACG GGCAGAAGGC TTAAGGTTTT TCGCCATTTC GACAACGTCG TTATCATAGA
[1495] 9841 AATAAAGCCC GGTTACCGCA TAATTACTTT TTGGTTGTAG CGGTTTTTCT TCCAGGCTTA
[1496] 9901 TTGCCGTACC GTTTTTATCA AACTCAACGA CGCCGTAGCG TTCAGGATCA TTAACGTGAT
[1497] 9961 AGGCAAATAC CGTTGCACCA CTTTCTTTGT TAACAGCGAC ATCCATTAAC TTCGGCAGAT
[1498] 10021 CATGACCGTA GAAGATATTA TCACCAAGAA CCAAAGCACA ATCATCACCA CCGATAAACT
[1499] 10081 CTTACCCGAT AATAAACGCC TGCGCAAGCC CATCTGGAGT CGGTTGCACT TTGTACTGAA
[1500] 10141 GATTTAGCCC CCACTGGCTA CCGTCACCTA GCAGTTGTTG AAAACGAGGA GTATCCTGTG
[1501] 10201 GCGTACTAAT AATCAGAATA TCGCGAATAC CCGCCAACAT CAGTGTAGAG AGCGGGTAAT
[1502] 10261 AGATCATCGG CTTATCATAA ATAGGTAATA GCTGTTTACT GACAGCCATA GTCACAGGAT
[1503] 10321 AAAGACGTGT ACCAGAACCA CCCGCTAAAA TAATACCTTT ACGCGTTTTT ATTTTCATCAT
[1504] 10381 TCCTTTTAAT TCATCTTGCT CCACCATCAC GAACAAGATG CAAAACTAT TAAATTGCTG
[1505] 10441 TAGTCGTAAT TAATTCGTTG AGCATTCGTT TCACACCAAC CTGCCAGTCA GGCAAGACAA
[1506] 10501 GCGCAAAGTT CTGCTGAAAT TTTTCTGTAT TAAGGCGAGA GTTATGTGGA CGACGAGCTG
[1507] 10561 GTGTAGGATA GGCTGTTGTT GGTACTGCGT TGAGCTTGTT GAGTGCAAGG GGAATACCTG
[1508] 10621 CTTTGC GCGC CTCTTCAAAA ACCAGCGCAG CATAATCGTG CCAGGTTGTG GTACCACTGG
[1509] 10681 CTACCAGATG GTACAAACCT GCGACTTCCG GTTTATTTCAG TGCCACACGA ATAGCATGTG
[1510] 10741 CCGTACAATC AGCCAGCAGC TCAGCACCTG TTGGCGCACC AAATTGATCA TTTATCACAG
[1511] 10801 CCAGTTCTTC GCGCTCTTTT GCCAGACGCA ACATCGTTTT GCGGAAGTTA TTTCTTTAG
[1512] 10861 CTGCGTATAC CCAGCTGGTA CGGAAAATAA GATGCTTCGC GCAATGTTCC TGTAACGCTT
[1513] 10921 TTTCTCCGGC TAACCTGGTT TCACCGTAAA CATTAGCGG TGCGGTTGCA TCCGTCTCCA
[1514] 10981 GCCATGGCGT GTCGCCATTT CCAGGGAATA CGTAGTCAGT TGAGTAATGA ATTACCAAG
[1515] 11041 CCCCAACTTC ATTAGCCTCT TTTGCAATTG ATTCAACACT AGTCGCATTG AGTAATTGTG
[1516] 11101 CAAATTTCGGG TTCTGACTCA GCCTTATCTA CTGCGGTGTG AGCCGCAGCA TTAACAATAA
[1517] 11161 CATCAGGTCG AATTCTTTTG ACTGTTTCAG CTACACCTTC AGGATTACTA AAATCACCAC
[1518] 11221 AATAATCAGT GGAGTGAACA TCAAGAGCAA TCAAATTACC CAAAGGTGCC AGAGCACGCT
[1519] 11281 GTAGTTCCCA ACCTACCTGC CCTGTTTTGC CGAAAAGGAG GATATTCATT ACTGGCGGCC
[1520] 11341 CTCATAGTTC TGTTCATCC ACGATTGATA AGCACCCTT TTCACATTAT CAACCCATTT
[1521] 11401 TGTATTGGAC AGGTACCATT CCAATGTCTT CCGAATCCCG CTCTCAAACG TTTCTGCGG
[1522] 11461 TTTCCAGCCC AATTCGCGG TAATCTTCTC TGCATCAATC GCATAACGGC GATCGTGTCC
[1523] 11521 CGGGCGATCG GCAACATAAG TAATTTGCTC GCGGTAAGAT TTCTCTTTCG GTACAATCTC
[1524] 11581 ATCCAGCAAA TCACAAATAG TGAGCACTAC ATCGATGTTT TTCTTTTCGT TGTGTCCACC
[1525] 11641 AATGTTATAA GTTTCACCCG CTTTACCTTC GGTACGACG GTATATAACG CACGCGCATG
[1526] 11701 ATCTTCAACA TACAGCCAGT CACGAATTTG ATCCCCTTTG CCATAAATAG GTAATGCCTT
[1527] 11761 ACCTTCCAGA GCATTCAGAA TAACCAATGG AATCAATTTT TCCGGGAAAT GATAAGGACC
[1528] 11821 ATAATTATTA GAGCAATTAG TCACAATGGT TGGTAAACCA TAGGTACGTT TCCACGCGCG
[1529] 11881 GACTAAATGA TCGCTGGATG CTTTTGAAGC GGAATAAGGG CTGCTTGGCG CGTAAGCTGT

[1530] 11941 TGTCTCTGTA AATAAGGGTA ATTCTTCTGT ATTATTTACC TCGTCAGGAT GAGGCAAATC
[1531] 12001 ACCATAGACT TCGTCAGTAG AAATATGATG AAAACGGAAT CTAGTTTTCT TGTCGCTATC
[1532] 12061 AAGAGCAGAC CAATAATTGC GAGCGGCTTC CAAAAGGACA TATGTACCAA CAATATTGGT
[1533] 12121 TTCAATAAAT GCCGCAGGAC CTGTAATTGA ACGGTCAACA TGGCTTTCAG CAGCCAGGTG
[1534] 12181 CATCACTGCA TCTGGCTGAT GCTGAGCAAA AATCCGTGCC ATTGCAGCTG CATCGCAAAT
[1535] 12241 ATCCGCATGT TCAAAAACAT AGCGTTCAGA ATCAGAAACA TCAGCAAGTG ATTCCAGGTT
[1536] 12301 TCCGGCGTAC GTTAATTTAT CGACATTAAC AACACTATCC TGCCTATTAT TTATAATGTG
[1537] 12361 ACGAACTACA GCAAAACCAA TAAATCCTGC GCCACCAGTA ACAAGTATTT TCACCTAATT
[1538] 12421 TATTCCATAT TGCTTCAGAG CATGCTGTGA AATAAGCGGC TCTCAGTTTG ATTAATAGAA
[1539] 12481 GTATTAATGC ACGCTACCGC CCCTGGCTTT ACAGCTACCA GAGCACTGCA TGCATGCCTA
[1540] 12541 CGATGTGACG AGCGTTACCC ACTCGCGCTA AACCCGAAAA ATTCAAAAGC TAATTGTCTT
[1541] 12601 ACCAATCCGC TCTGGAAACA AGGAAAATCC TGGAAAACCT TGACTAAAAT CCTATTGCTA
[1542] 12661 ACTCGTTGTT ATTCTGATTG TTTATATAAA ACAACGGCAG GAATATTCGC AACAAATTAC
[1543] 12721 TTTCACCACG AATCTTCACT GCCGTTATAA TTTTCTTATC AACCGTTACA TCCGGTCAGA
[1544] 12781 TTTTCATTAT TCGCTTAACA GCTTCTCAAT ACCTTTACGG AACTTCGCCC CTTCCTTCAG
[1545] 12841 GTTGC GCAGC CCATACTTCA CAAACGCCTG CATATAGCCC ATTTTTTTAC CGCAGTCGTA
[1546] 12901 GCTGTCGCCC GTCATCAGCA TTGCATCAAC GGACTGTTTT TTCGCCAGCT CGGCAATGGC
[1547] 12961 ATCAGTCAGC TGAATACGTC CCCATGCACC AGGCTGAGTA CGTTCAAGTT CCGGCCAAAT
[1548] 13021 ATCGGCAGAA AGCACATAGC GACCAACGGC CATGATGTCT GAGTCCAGCG TCTGCGGCTG
[1549] 13081 ATCCGGTTTT TCGATAAATT CAACAATGCG GCTGACTTTA CCTTCGCGAT CCAGCGGTTT
[1550] 13141 TTTGGTCTGG ATGACGGAGT ATTCAGAGAG GTCACCCGGC ATACGTTTTG CCAGCACCTG
[1551] 13201 GCTACGGCCC GTTTCATTGA AGCGCGCAAT CATGGCAGCA AGGTTGTAGC GTAGCGGGTC
[1552] 13261 GGCGCTGGCG TCGTCGATCA CAACGTCTGG CAGCACCACG ACAAATGGAT TGTCACCAAT
[1553] 13321 GGCGGGTCGT GCACACAAAA TGGAGTGACC TAAACCTAAA GGTTCCGCCCT GACGCACGTT
[1554] 13381 CATAATAGTC ACGCCCGGCG GGCAGATAGA TTGCACTTCC GCCAGTAGTT GACGCTTCAC
[1555] 13441 GCGCTGCTCA AGGAGAGATT CTAATTCATA AGAGGTGTGCG AAGTGGTTTT CGACCGCGTT
[1556] 13501 CTTGGACGCA TGAGTTACCA GGAGGATTTT TTTGATCCCT GCAGCCACAA TCTCGTCAAC
[1557] 13561 AATGTACTGA ATCATTTGGT TGTCGACGAT CGGTAGCATC TCTTTGGGTA TCGCCTTAGT
[1558] 13621 GGCAGGCAAC ATATGCATCC CAAGACCGC TACCGGTATA ACTGCTTTTA AATTCGTCAT
[1559] 13681 TATTTTCCTA CCTCTAAGGG GCTGATAGTG CGTAAATTAT TGTCATAGGT TAGCCAAACG
[1560] 13741 GTATGGCTAT ATACCAAGCA TAACTTTGAT TAAACCTTAC GATAACACTA CACACCATCA
[1561] 13801 GCATCTGGGT TACTCGGATT ACTCGGAAAT CCACATACTG ATAATTTAAT CAGTACCTCT
[1562] 13861 TTCCGAATAA TCGTAGTCCA ACCTGGTCCT TTTTCTCTG ACTCGTCTGC ATTACTCAGA
[1563] 13921 AACAAACGTT ATGTCGTCTT TTTTGGCATG GACGAATTCA TACTGCAGAG TTCGATCCAG
[1564] 13981 ACCTTGCGAC AGCGTATACG GTGCAACAAA ACCTGAAGAA TGCACTTTTCG TTGCGTCAAA
[1565] 14041 CTGTGTTGTT GCGCAGAATT TTTTCACGCG CACAGAGCTG ACAGCGTATT TTTTGCCCGT
[1566] 14101 AATTTTGCTC AGGATATCAA AGCAATATCC ACCCAGCATT CCTAGTGGGT AAGGCAAGTG
[1567] 14161 CATAGAAGGG ATCTTTTTGT TCAGGCTTTG TTCAACTTCA GCAACCAACT GGTTCATGTT
[1568] 14221 CAGGTCTGGC TTATCAACAT AGTTATAAAC CTCATAACCT GCGGCAACAT TCTTCAGTTT

[1569] 14281 G T A C T T G A T A A A C T C A A C A A T G T T T C C A A C A T A A G C C A T G G A C T T A T A G T T A G T C C C T G C
[1570] 14341 G C C C A C C A T C A T A A A A C T T G C C G C C A G C G A T C T G T T T C A G C A A G T T A T A G A C G T T A C C G C G
[1571] 14401 G T T G C G T T C A C C G A A G A T A A C G G T A G G A C G G A T G A T G G T T A A T G A A C G T T C T G T T G G T G C
[1572] 14461 T T T G T T A T A C C A T T C A C G C A G C A C T T C C T C T G C C T G C C A C T T A C T T T T G C C G T A G T G G T T
[1573] 14521 G A A A G G G T C G T G T G G A T G G T T T T C G T C A G G G T T G T G T G T T T G T T C A A A C C A T A A A C A G C A A C
[1574] 14581 G G A A C T G G T A A A G A T G A T A T T T T A A C G C C A T T T T T T T C C A T G G C C G C C A G C A C A T T G C G
[1575] 14641 G G T A C C C T G A A C G T T G A C A T C A T A A T A G A G A G A A G T A G G G C T G A C G T C A T C G C G G T G T T C
[1576] 14701 C G C T G C C A G T A G T A C A A C A G T G T C A A A A C C G G C T A A C G C C T G G T C G A G T G C C T G T T G A T C
[1577] 14761 A C G A A C A T C A C C A A T C T G T G T G A T T T C T G G A T A A A A G T G G C T C T G C C G T T T G T C C A G G T T
[1578] 14821 C T T G A T A T T A A A G T C A G C A A T T G C C G T T T C A A G T A G T C G G G T T C C T A C G A A T C C G G A A G C
[1579] 14881 T C C T A T G A G C A A A A C G T T A T T G T T C A T A A A T C A C T T T A G T C T G G T T G T T A C G T A A G A A A C
[1580] 14941 A C A A G A T A A A G A T G A G T A C C T T C C C T G A G T A G T C A A T G C T G C C C A G C C C C A G C T T T A A C A
[1581] 15001 G T T A G T G T G A G G A T T A T A A T C T T T T A G A A C A T T A T A T C C A G T A A G T T T A T G A A T G G T C G C
[1582] 15061 A A A T C T A C T C T C T C C G T T C C G G C A A T C T A A A G T T A A T G C T A G C G A C G T C G T G G G A T C C T C
[1583] 15121 T A G A G T C G A C C T G C A G G C A T G C A A G C T T G A G T A T T C T A T A G T C T A C C T A A A T A G C T T G G
[1584] 15181 C G T A A T C A T G G T C A T A G C T G T T T C C T G T G T G A A A T T G T T A T C C G C T C A C A T T C C A C A C A
[1585] 15241 A C A T A C G A G C C G G A A G C A T A A G T G T A A A G C C T G G G G T G C C T A A T G A G T G A G C T A A C T C A
[1586] 15301 C A T T A A T T G C G T T G C G T C A C T G C C C G C T T T C C A G T C G G G A A A C C T G T C G T G C C A G C T G C
[1587] 15361 A T T A A T G A A T C G G C C A A C G C G A A C C C C T T G C G G C C G C C C G G C C G T C G A C A A T T C T C A T
[1588] 15421 G T T T G A C A G C T T A T C A T C G A A T T T C T G C C A T T C A T C C G C T A T T A T C A C T A T T C A G G C G
[1589] 15481 T A G C A A C C A G G C G T T T A A G G G C A C C A A T A A C T G C C T T A A A A A A T T A C G C C C G C C C T G C
[1590] 15541 C A C T C A T C G C A G T A C T G T T G T A A T T C A T T A A G C A T T C T G C C G A C A T G G A A G C C A T C A C A A
[1591] 15601 A C G G C A T G A T G A A C C T G A A T C G C C A G C G G C A T C A G C A C C T T G T C G C C C T T G C G T A T A A T A T
[1592] 15661 T T G C C C A T G G T G A A A A C G G G G G C G A A G A A G T T G T C C A T A T T G G C C A C G T T T A A A T C A A A A
[1593] 15721 C T G G T G A A A C T C A C C C A G G G A T T G G C T G A G A C G A A A A A C A T A T T C T C A A T A A A C C C T T T A
[1594] 15781 G G G A A T A G G C C A G G T T T T C A C C G T A A C A C G C C A C A T C T T G C G A A T A T A T G T G T A G A A A C
[1595] 15841 T G C C G G A A A T C G T C G T G G T A T T C A C T C C A G A G C G A T G A A A A C G T T T C A G T T T G C T A T G G
[1596] 15901 A A A A C G G T G T A A C A A G G G T G A A C A C T A T C C A T A T C A C C A G C T C A C C G T C T T T C A T T G C C
[1597] 15961 A T A C G A A A T T C C G G A T G A G C A T T C A T C A G G C G G G C A A G A A T G T G A A T A A A G G C C G A T A A
[1598] 16021 A A C T T G T G C T T A T T T T T C T T T A C G G T C T T T A A A A G G C C G T A A T A T C C A G C T G A A C G G T C
[1599] 16081 T G G T T A T A G G T A C A T T G A G C A A C T G A C T G A A A T G C C T C A A A A T G T T C T T T A C G A T G C C A T
[1600] 16141 T G G G A T A T A T C A A C G G T G G T A T A T C C A G T G A T T T T T T T C T C C A T T T T A G C T T C C T T A G C T
[1601] 16201 C C T G A A A A T C T C G A T A A C T C A A A A A A T A C G C C C G G T A G T G A T C T A T T T T C A T T A T G G T G A
[1602] 16261 A A G T T G G A A C C T C T T A C G T G C C G A T C A A C G T C T C A T T T T C G C C A A A A G T T G G C C A G G G C
[1603] 16321 T T C C C G G T A T C A A C A G G G A C A C C A G G A T T T A T T A T T C T G C G A A G T G A T C T T C C G T C A C A
[1604] 16381 G G T A T T T A T T C G C G A T A A G C T C A T G G A G C G G C G T A A C C G T C G C A C A G G A A G G A C A G A G A A
[1605] 16441 A G C G C G G A T C T G G G A A G T G A C G G A C A G A A C G G T C A G G A C C T G G A T T G G G G A G G C G G T T G C
[1606] 16501 C G C C G C T G C T G C T G A C G G T G T G A C G T T C T C T G T T C C G G T C A C A C C A C A T A C G T T C C G C C A
[1607] 16561 T T C C T A T G C G A T G C A C A T G C T G T A T G C C G G T A T A C C G C T G A A A G T T C T G C A A A G C C T G A T

[1608] 16621 GGGACATAAG TCCATCAGTT CAACGGAAGT CTACACGAAG GTTTTTGCGC TGGATGTGGC
[1609] 16681 TGCCCGGCAC CGGGTGCAGT TTGCGATGCC GGAGTCTGAT GCGGTTGCGA TGCTGAAACA
[1610] 16741 ATTATCCTGA GAATAAATGC CTTGGCCTTT ATATGGAAAT GTGGAAGTGA GTGGATATGC
[1611] 16801 TGTTTTTGTC TGTTAAACAG AGAAGCTGGC TGTTATCCAC TGAGAAGCGA ACGAAACAGT
[1612] 16861 CGGGAAAATC TCCCATTATC GTAGAGATCC GCATTATTAA TCTCAGGAGC CTGTGTAGCG
[1613] 16921 TTTATAGGAA GTAGTGTCTT GTCATGATGC CTGCAAGCGG TAACGAAAAC GATTTGAATA
[1614] 16981 TGCCTTCAGG AACAAATAGAA ATCTTCGTGC GGTGTTACGT TGAAGTGGAG CGGATTATGT
[1615] 17041 CAGCAATGGA CAGAACAACC TAATGAACAC AGAACCATGA TGTGGTCTGT CCTTTTACAG
[1616] 17101 CCAGTAGTGC TCGCCGAGT CGAGCGACAG GGCGAAGCCC TCGGCTGGTT GCCCTCGCCG
[1617] 17161 CTGGGCTGGC GGCCGTCTAT GGCCCTGCAA ACGCGCCAGA AACGCCGTCG AAGCCGTGTG
[1618] 17221 CGAGACACCG CGGCCGGCCG CCGGCGTTGT GGATACCTCG CGGAAAACCTT GGCCCTCACT
[1619] 17281 GACAGATGAG GGGCGGACGT TGACACTTGA GGGGCCGACT CACCCGGCGC GGC GTTGACA
[1620] 17341 GATGAGGGGC AGGCTCGATT TCGGCCGGCG ACGTGGAGCT GGCCAGCCTC GCAAATCGGC
[1621] 17401 GAAAACGCCT GATTTTACGC GAGTTTCCCA CAGATGATGT GGACAAGCCT GGGGATAAGT
[1622] 17461 GCCCTGCGGT ATTGACACTT GAGGGGCGCG ACTACTGACA GATGAGGGGC GCGATCCTTG
[1623] 17521 ACACTTGAGG GGCAGAGTGC TGACAGATGA GGGGCGCACC TATTGACATT TGAGGGGCTG
[1624] 17581 TCCACAGGCA GAAAATCCAG CATTTGCAAG GGTTTCCGCC CGTTTTTCGG CCACCGCTAA
[1625] 17641 CCTGTCTTTT AACCTGCTTT TAAACCAATA TTTATAAACC TTGTTTTTAA CCAGGGCTGC
[1626] 17701 GCCCTGTGCG CGTGACCGCG CACGCCGAAG GGGGGTGCCC CCCCTTCTCG AACCCCTCCG
[1627] 17761 GTCGAGTGAG CGAGGAAGCA CCAGGGAACA GCACTTATAT ATTCTGCTTA CACACGATGC
[1628] 17821 CTGAAAAAAC TTCCCTTGGG GTTATCCACT TATCCACGGG GATATTTTTA TAATTATTTT
[1629] 17881 TTTTATAGTT TTTAGATCTT CTTTTTTAGA GCGCCTTGTA GGCCTTTATC CATGCTGGTT
[1630] 17941 CTAGAGAAGG TGTTGTGACA AATTGCCCTT TCAGTGTGAC AAATCACCTT CAAATGACAG
[1631] 18001 TCCTGTCTGT GACAAATTGC CCTTAACCCT GTGACAAATT GCCCTCAGAA GAAGCTGTTT
[1632] 18061 TTTCACAAAG TTATCCCTGC TTATTGACTC TTTTTTATTT AGTGTGACAA TCTAAAACT
[1633] 18121 TGTCACACTT CACATGGATC TGTCATGGCG GAAACAGCGG TTATCAATCA CAAGAAACGT
[1634] 18181 AAAAATAGCC CGCGAATCGT CCAGTCAAAC GACCTCACTG AGGCGGCATA TAGTCTCTCC
[1635] 18241 CGGGATCAAA AACGTATGCT GTATCTGTTT GTTGACCAGA TCAGAAAATC TGATGGCACC
[1636] 18301 CTACAGGAAC ATGACGGTAT CTGCGAGATC CATGTTGCTA AATATGCTGA AATATTCGGA
[1637] 18361 TTGACCTCTG CGGAAGCCAG TAAGGATATA CGGCAGGCAT TGAAGAGTTT CGCGGGGAAG
[1638] 18421 GAAGTGGTTT TTTATCGCCC TGAAGAGGAT GCCGGCGATG AAAAAGGCTA TGAATCTTTT
[1639] 18481 CCTTGGTTTA TCAAACGTGC GCACAGTCCA TCCAGAGGGC TTTACAGTGT ACATATCAAC
[1640] 18541 CCATATCTCA TTCCCTTCTT TATCGGGTTA CAGAACCGGT TTACGCAGTT TCGGCTTAGT
[1641] 18601 GAAACAAAAG AAATCACCAA TCCGTATGCC ATGCGTTTAT ACGAATCCCT GTGTCAGTAT
[1642] 18661 CGTAAGCCGG ATGGCTCAGG CATCGTCTCT CTGAAAATCG ACTGGATCAT AGAGCGTTAC
[1643] 18721 CAGCTGCCTC AAAGTTACCA GCGTATGCCT GACTTCCGCC GCCGTTCTCT GCAGGTCTGT
[1644] 18781 GTTAATGAGA TCAACAGCAG AACTCCAATG CGCCTCTCAT ACATTGAGAA AAAGAAAGGC
[1645] 18841 CGCCAGACGA CTCATATCGT ATTTTCCTTC CGCGATATCA CTTCCATGAC GACAGGATAG
[1646] 18901 TCTGAGGGTT ATCTGTCACA GATTTGAGGG TGGTTCGTCA CATTTGTTCT GACCTACTGA

[1647] 18961 GGGTAATTTG TCACAGTTTT GCTGTTTCCT TCAGCCTGCA TGGATTTTCT CATACTTTTT
[1648] 19021 GAACTGTAAT TTTTAAGGAA GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCTTTCT
[1649] 19081 CTTTCCCTTC GTCATGTGAC CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG
[1650] 19141 ATTATCACAG TTTATTACTC TGAATTGGCT ATCCGCGTGT GTACCTCTAC CTGGAGTTTT
[1651] 19201 TCCCACGGTG GATATTTCTT CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT
[1652] 19261 CTTTGCTTCC TCGCCAGTTC GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT
[1653] 19321 GATAATAAGT GACTGAGGTA TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT
[1654] 19381 TAAACAACCT TGC GGTTTTT TGATGACTTT GCGATTTTGT TGTTGCTTTG CAGTAAATTG
[1655] 19441 CAAGATTTAA TAAAAAACG CAAAGCAATG ATTAAAGGAT GTTCAGAATG AAACATCATG
[1656] 19501 AAACACTTAA CCAGTGCATA AACGCTGGTC ATGAAATGAC GAAGGCTATC GCCATTGCAC
[1657] 19561 AGTTTAATGA TGACAGCCCG GAAGCGAGGA AAATAACCCG GCGCTGGAGA ATAGGTGAAG
[1658] 19621 CAGCGGATTT AGTTGGGGTT TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC
[1659] 19681 TACCGCACCC GGATATGGAA ATTCGAGGAC GGGTTGAGCA ACGTGTGGT TATACAATTG
[1660] 19741 AACAAATTA TCATATGCGT GATGTGTTTG GTACGCGATT GCGACGTGCT GAAGACGTAT
[1661] 19801 TTCCACCGGT GATCGGGGTT GCTGCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG
[1662] 19861 TTCATCTTGC TCAGGATCTG GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG
[1663] 19921 ACCCCCAGGG AACAGCCTCA ATGTATCACG GATGGGTACC AGATCTTCAT ATTCATGCAG
[1664] 19981 AAGACACTCT CCTGCCTTTC TATCTTGGGG AAAAGGACGA TGTCACCTAT GCAATAAAGC
[1665] 20041 CCACTTGCTG GCCGGGGCTT GACATTATTC CTTCCTGTCT GGCTCTGCAC CGTATTGAAA
[1666] 20101 CTGAGTTAAT GGGCAAATTT GATGAAGGTA AACTGCCCAC CGATCCACAC CTGATGCTCC
[1667] 20161 GACTGGCCAT TGAAACTGTT GCTCATGACT ATGATGTCAT AGTTATTGAC AGCGCGCCTA
[1668] 20221 ACCTGGGTAT CGGCACGATT AATGTCGTAT GTGCTGCTGA TGTGCTGATT GTTCCCACGC
[1669] 20281 CTGCTGAGTT GTTTGACTAC ACCTCCGCAC TGCAGTTTTT CGATATGCTT CGTGATCTGC
[1670] 20341 TCAAGAACGT TGATCTTAAA GGGTTCGAGC CTGATGTACG TATTTTGCTT ACCAAATACA
[1671] 20401 GCAATAGTAA TGGCTCTCAG TCCCCGTGGA TGGAGGAGCA AATTCGGGAT GCCTGGGGAA
[1672] 20461 GCATGGTTCT AAAAAATGTT GTACGTGAAA CGGATGAAGT TGGTAAAGGT CAGATCCGGA
[1673] 20521 TGAGAACTGT TTTTGAACAG GCCATTGATC AACGCTCTTC AACTGGTGCC TGGAGAAATG
[1674] 20581 CTCTTTCTAT TTGGGAACCT GTCTGCAATG AAATTTTCGA TCGTCTGATT AAACCACGCT
[1675] 20641 GGGAGATTAG ATAATGAAGC GTGCGCCTGT TATTCAAAA CATACGCTCA ATACTCAACC
[1676] 20701 GGTTGAAGAT ACTTCGTTAT CGACACCAGC TGCCCCGATG GTGGATTCTG TAATTGCGCG
[1677] 20761 CGTAGGAGTA ATGGCTCGCG GTAATGCCAT TACTTTGCCT GTATGTGGTC GGGATGTGAA
[1678] 20821 GTTTACTCTT GAAGTGCTCC GGGGTGATAG TGTGAGAAG ACCTCTCGGG TATGGTCAGG
[1679] 20881 TAATGAACGT GACCAGGAGC TGCTTACTGA GGACGCACTG GATGATCTCA TCCCTTCTTT
[1680] 20941 TCTACTGACT GGTCAACAGA CACCGGCGTT CGGTCAAGA GTATCTGGTG TCATAGAAAT
[1681] 21001 TGCCGATGGG AGTCGCCGTC GTAAAGCTGC TGCACTTACC GAAAGTGATT ATCGTGTCTT
[1682] 21061 GGTTGGCGAG CTGGATGATG AGCAGATGGC TGCATTATCC AGATTGGGTA ACGATTATCG
[1683] 21121 CCCAACAAGT GCTTATGAAC GTGGTCAGCG TTATGCAAGC CGATTGCAGA ATGAATTTGC
[1684] 21181 TGGAATATT TCTGCGCTGG CTGATGCGGA AAATATTTCA CGTAAGATTA TTACCCGCTG
[1685] 21241 TATCAACACC GCCAAATTGC CTAAATCAGT TGTGCTCTT TTTTCTCACC CCGGTGAACT

[1686] 21301 ATCTGCCCCG TCAGGTGATG CACTTCAAAA AGCCTTTACA GATAAAGAGG AATTACTTAA
[1687] 21361 GCAGCAGGCA TCTAACCTTC ATGAGCAGAA AAAAGCTGGG GTGATATTTG AAGCTGAAGA
[1688] 21421 AGTTATCACT CTTTTAACTT CTGTGCTTAA AACGTCATCT GCATCAAGAA CTAGTTTAAG
[1689] 21481 CTCACGACAT CAGTTTGCTC CTGGAGCGAC AGTATTGTAT AAGGGCGATA AAATGGTGCT
[1690] 21541 TAACCTGGAC AGGTCTCGTG TTCCAAGTGA GTGTATAGAG AAAATTGAGG CCATTCTTAA
[1691] 21601 GGAACCTGAA AAGCCAGCAC CCTGATGCGA CCACGTTTTA GTCTACGTTT ATCTGTCTTT
[1692] 21661 ACTTAATGTC CTTTGTTACA GGCCAGAAAG CATAACTGGC CTGAATATTC TCTCTGGGCC
[1693] 21721 CACTGTTCCA CTTGTATCGT CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT
[1694] 21781 CGTCGGTCTG ATTATTAGTC TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA
[1695] 21841 GTCTGGGACC ACGGTCCAC TCGTATCGTC GGTCTGATAA TCAGACTGGG ACCACGGTCC
[1696] 21901 CACTCGTATC GTCGGTCTGA TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC
[1697] 21961 TGATTATTAG TCTGGGACCA CGGTCCCACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA
[1698] 22021 CCACGGTCCC ACTCGTATCG TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA
[1699] 22081 TCGTCGGTCT GATTATTAGT CTGGGACCAC GATCCCACTC GTGTTGTCTG TCTGATTATC
[1700] 22141 GGTCTGGGAC CACGGTCCCA CTTGTATTGT CGATCAGACT ATCAGCGTGA GACTACGATT
[1701] 22201 CCATCAATGC CTGTCAAGGG CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA
[1702] 22261 CCTCGGTGTG CGGTTGTATG CCTGCTGTGG ATTGCTGCTG TGTCTGCTT ATCCACAACA
[1703] 22321 TTTTGCGCAC GGTATGTGG AAAAAATACC TGGTTACCCA GGCCGTGCCG GCACGTTAAC
[1704] 22381 CGGGCTGCAT CCGATGCAAG TGTGTCGCTG TCGACGAGCT CGCGAGCTCG GACATGAGGT
[1705] 22441 TGCCCCGTAT TCAGTGTCGC TGATTTGTAT TGTCTGAAGT TGTTTTTACG TTAAGTTGAT
[1706] 22501 GCAGATCAAT TAATACGATA CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT
[1707] 22561 CCACGCACGT TGTGATATGT AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA
[1708] 22621 TCCGACAGGT TACGGGGCGG CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT
[1709] 22681 TTAAGGCGTT TCCGTTCTTC TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA
[1710] 22741 AAAGAAAGGA AACGACAGGT GCTGAAAGCG AGCTTTTTTG CCTCTGTCGT TTCCTTTCTC
[1711] 22801 TGTTTTTGTC CGTGGAATGA ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG
[1712] 22861 CACACATTAT ACGAAGTTAT ATTCGAT

序列表

<110> 格林考瓦因有限公司

<120> 在原核细胞中产生免疫原性多糖的生物合成系统

<130> 031229.0005-W001

<140> PCT/US10/02980

<141> 2010-11-16

<150> 61/272,931

<151> 2009-11-19

<160> 30

<170> PatentIn version 3.5

<210> 1

<211> 993

<212> DNA

<213> 大肠杆菌

<400> 1

```

atgaacgata acgttttgc tcataggagct tccgattcg taggaaccgc actacttgaa      60
acggcaattg ctgacttta tatcaagaac ctggacaaac agcagagcca cttttatcca      120
gaaatcacac agattggcga tgttcggat caacaggcac tcgaccaggc gttagtgggt      180
tttgacaactg ttgtactact ggcagcggaa caccgggatg acgtcagccc tactctctc      240
tattatgatg tcaacgttca gggtaaccgc aatgtgctgg cgcccatgga aaaaaatggc      300
gttaaaaata tcatttttac cagttccgtt gctgtttatg gtttgaacaa acacaacct      360
gacgaaaacc atccacaaga ccctttcaac cactacggca aaagtaagtg gcagcagag      420
gaagtgcgc gtgaatggta taacaaagca ccaacagAAC gtccattaac catcatcgt      480
ctaccgtta tcttcggta acgcaaccgc ggtaacgtct ataacttgc taaacagatc      540
gtcggcgga agtttatgat ggtggcgga gggactaac ataagtcac ggttatggt      600
ggaaacattg ttgagtttat caagtacaaa ctgaagaatg ttgcgcagg ttatgaggt      660
tataactacg ttgataagcc agacctgaac atgaaccagt tggttgctga agttgaacaa      720
agcctgaaca aaaagatccc ttctatgac ttgccttacc cactaggaat gctgggtgga      780
tattgctttg atacctgag caaaattacg ggcaaaaaat acgtgtcag ctcagtgcgc      840
gtgaaaaaat tctgcgcaac aacacagttt gacgcaacga aagtgcattc ttacgtttt      900
gtggcaccgt atacgtgtc gcaaggctg gategaacac tgcagtatga attcgttcac      960
gccaaaaaag acgacataac gtttgtttct gag      993

```

<210> 2

<211> 331

<212> PRT

<213> 大肠杆菌

<400> 2

```

Met Asn Asp Asn Val Leu Leu Ile Gly Ala Ser Gly Phe Val Gly Thr      1
1           5           10          15

Arg Leu Leu Glu Thr Ala Ile Ala Asp Phe Asn Ile Lys Asn Leu Asp      20
20          25          30

Lys Gln Gln Ser His Phe Tyr Pro Glu Ile Thr Gln Ile Gly Asp Val      35
35          40          45

```

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

[0002]

<210> 3
 <211> 993
 <212> DNA
 <213> 大肠杆菌

<400> 3
 atgaacgata acgttttget cataggaget tccggattcg taggaacecg actacttgaa 60
 acggcaattg ctgaatttaa tatcaagaac ctggacaaac agcagagcca ettttatcca 120
 gaaatcacac agattgggtga tgttctgat caacaggcac tcgaccaggc gttagccggt 180
 tttagacactg ttgtgtact ggcageggaa caccgcgatg acgtcagccc tacttctetc 240
 tatttatgatg tcaacgttca ggttaccgc aatgtgctgg cgcccatgga aaaaaatggc 300
 gttaaaaata tcatctttac cagttccgtt gctgtttaig gtttgaacaa acacaacct 360
 gacgaaaacc atccacacga tctttcaac cactacggca aaagtaagtg gcaggcagag 420
 gaagtgtgc gtgaatggta taacaaagca ccaacagaac gttcattaac catcatccgt 480
 cctaccgtta tcttcgggtga acggaaccgc ggtaacgtct ataacttget gaaacagatc 540
 gctggcgcca agtttatgat ggtggcgcca gggactaact ataagtcacat ggcttatggt 600
 ggaaacattg ttgagtttat caagtacaaa ctgaagaatg ttgccgcagg ttacgaggtt 660
 tataactacg ttgataagcc agacctgaac atgaaccagt tggttgtga agttgaacaa 720
 agcctgaaca aaaagatccc ttctatgcac ttgccttacc cactaggaat gctgggtgga 780
 tatgtcttg atatcctgag caaaattacg ggcaaaaaat acgtgtcag ctcgtgtgcg 840
 gtgaaaaat tctgcgaac aacacagttt gacgcaacga aagtgcattc ttcaggtttt 900
 gtggcaccgt atacgtgtc gcaaggctg gatcgaactc tgcagtatga attcgtccat 960
 gccaaaaag acgacataac gttgtttct gag 993

[0003]

<210> 4
 <211> 331
 <212> PRT
 <213> 大肠杆菌

<400> 4
 Met Asn Asp Asn Val Leu Leu Ile Gly Ala Ser Gly Phe Val Gly Thr
 1 5 10 15
 Arg Leu Leu Glu Thr Ala Ile Ala Asp Phe Asn Ile Lys Asn Leu Asp
 20 25 30
 Lys Gln Gln Ser His Phe Tyr Pro Glu Ile Thr Gln Ile Gly Asp Val
 35 40 45
 Arg Asp Gln Gln Ala Leu Asp Gln Ala Leu Ala Gly Phe Asp Thr Val
 50 55 60
 Val Leu Leu Ala Ala Glu His Arg Asp Asp Val Ser Pro Thr Ser Leu
 65 70 75 80
 Tyr Tyr Asp Val Asn Val Gln Gly Thr Arg Asn Val Leu Ala Ala Met
 85 90 95
 Glu Lys Asn Gly Val Lys Asn Ile Ile Phe Thr Ser Ser Val Ala Val
 100 105 110
 Tyr Gly Leu Asn Lys His Asn Pro Asp Glu Asn His Pro His Asp Pro
 115 120 125
 Phe Asn His Tyr Gly Lys Ser Lys Trp Gln Ala Glu Glu Val Leu Arg

130	135	140
Glu Trp Tyr Asn Lys Ala Pro Thr Glu Arg Ser Leu Thr Ile Ile Arg		
145	150	155 160
Pro Thr Val Ile Phe Gly Glu Arg Asn Arg Gly Asn Val Tyr Asn Leu		
165	170	175
Leu Lys Gln Ile Ala Gly Gly Lys Phe Met Met Val Gly Ala Gly Thr		
180	185	190
Asn Tyr Lys Ser Met Ala Tyr Val Gly Asn Ile Val Glu Phe Ile Lys		
195	200	205
Tyr Lys Leu Lys Asn Val Ala Ala Gly Tyr Glu Val Tyr Asn Tyr Val		
210	215	220
Asp Lys Pro Asp Leu Asn Met Asn Gln Leu Val Ala Glu Val Glu Gln		
225	230	235 240
Ser Leu Asn Lys Lys Ile Pro Ser Met His Leu Pro Tyr Pro Leu Gly		
245	250	255
Met Leu Gly Gly Tyr Cys Phe Asp Ile Leu Ser Lys Ile Thr Gly Lys		
260	265	270
Lys Tyr Ala Val Ser Ser Val Arg Val Lys Lys Phe Cys Ala Thr Thr		
275	280	285
Gln Phe Asp Ala Thr Lys Val His Ser Ser Gly Phe Val Ala Pro Tyr		
290	295	300
Thr Leu Ser Gln Gly Leu Asp Arg Thr Leu Gln Tyr Glu Phe Val His		
305	310	315 320
Ala Lys Lys Asp Asp Ile Thr Phe Val Ser Glu		
325	330	

[0004]

<210> 5
 <211> 993
 <212> DNA
 <213> 大肠杆菌

<400> 5
 atgaacgata acgttttgct cataggagct tccggattcg taggaacccg actacttgaa 60
 acggcaattg ctgaacttaa tatcaagaac ctggacaaac agcagagcca cttttateca 120
 gaaatcacac agattggiga tgttcgtgat caacaggcac tcgaccaggc gttagccggt 180
 tttagactg ttgtactact ggcagcggaa caccgcgaig acgtcagccc tacttctctc 240
 tattatgatg tcaacgttca ggtaccgcg aatgtgctgg cggccatgga aaaaaatggc 300
 gttaaaaata teatctttac cagttccgtt gctgtttatg gtttgaacaa acacaaccct 360
 gacgaaaacc atccacacga cctttcaac cactacggca aaagcaagtg gcagcgggag 420
 gaagtgtgc gtgaatggtg taacaaagca ccaacagaac gttcattaac tatcatcgt 480
 cctaccgtta tcttcggtga acgcaaccgc ggtaacgtct ataacttgcg gaaacagatc 540
 gctggcggca agtttatgat ggtgggcgca gggactaact ataagtccat ggcttatgtt 600

```

ggaacattg ttgagtttat caagtacaaa ctgaagaatg ttgccgcagg ttacgaggtt      660
tataactacg ttgataagcc agacetgaac atgaaccagt tggttgctga agttgaacaa      720
agcctgaaca aaaagatccc ttctatgcac ttgccttacc cactaggaat gctgggtgga      780
tattgctttg atatcctgag caaaattacg ggcaaaaaat acgctgtcag ctctgtgegc      840
gtgaaaaaat tctgcgcaac aacacagttt gacgcaacga aagtgcattc ttcaggtttt      900
gtggcaccgt atacgtgtc gcaaggctcg gatcgaaact tgcagtatga attcgcccat      960
gccaaaaaag acgacataac gtttgtttct gag                                  993

```

<210> 6
 <211> 331
 <212> PRT
 <213> 大肠杆菌

<400> 6
 Met Asn Asp Asn Val Leu Leu Ile Gly Ala Ser Gly Phe Val Gly Thr
 1 5 10 15

Arg Leu Leu Glu Thr Ala Ile Ala Asp Phe Asn Ile Lys Asn Leu Asp
 20 25 30

Lys Gln Gln Ser His Phe Tyr Pro Glu Ile Thr Gln Ile Gly Asp Val
 35 40 45

Arg Asp Gln Gln Ala Leu Asp Gln Ala Leu Ala Gly Phe Asp Thr Val
 50 55 60

[0005] Val Leu Leu Ala Ala Glu His Arg Asp Asp Val Ser Pro Thr Ser Leu
 65 70 75 80

Tyr Tyr Asp Val Asn Val Gln Gly Thr Arg Asn Val Leu Ala Ala Met
 85 90 95

Glu Lys Asn Gly Val Lys Asn Ile Ile Phe Thr Ser Ser Val Ala Val
 100 105 110

Tyr Gly Leu Asn Lys His Asn Pro Asp Glu Asn His Pro His Asp Pro
 115 120 125

Phe Asn His Tyr Gly Lys Ser Lys Trp Gln Ala Glu Glu Val Leu Arg
 130 135 140

Glu Trp Tyr Asn Lys Ala Pro Thr Glu Arg Ser Leu Thr Ile Ile Arg
 145 150 155 160

Pro Thr Val Ile Phe Gly Glu Arg Asn Arg Gly Asn Val Tyr Asn Leu
 165 170 175

Leu Lys Gln Ile Ala Gly Gly Lys Phe Met Met Val Gly Ala Gly Thr
 180 185 190

Asn Tyr Lys Ser Met Ala Tyr Val Gly Asn Ile Val Glu Phe Ile Lys
 195 200 205

Tyr Lys Leu Lys Asn Val Ala Ala Gly Tyr Glu Val Tyr Asn Tyr Val
 210 215 220

Asp Lys Pro Asp Leu Asn Met Asn Gln Leu Val Ala Glu Val Glu Gln
225 230 235 240

Ser Leu Asn Lys Lys Ile Pro Ser Met His Leu Pro Tyr Pro Leu Gly
245 250 255

Met Leu Gly Gly Tyr Cys Phe Asp Ile Leu Ser Lys Ile Thr Gly Lys
260 265 270

Lys Tyr Ala Val Ser Ser Val Arg Val Lys Lys Phe Cys Ala Thr Thr
275 280 285

Gln Phe Asp Ala Thr Lys Val His Ser Ser Gly Phe Val Ala Pro Tyr
290 295 300

Thr Leu Ser Gln Gly Leu Asp Arg Thr Leu Gln Tyr Glu Phe Val His
305 310 315 320

Ala Lys Lys Asp Asp Ile Thr Phe Val Ser Glu
325 330

<210> 7

<211> 993

<212> DNA

<213> 鲍氏志贺氏菌

<400> 7

[0006]

```
atgaacgata acgttttget cataggagct tccggattcg taggaacccg actacttgaa      60
acggcaattg ctgaacttta tatcaagaac ctggacaaac agcagagcca tttttatcca      120
gcaatcacac agattggcga tgttcgtgat caacaggeac tcgaccagge gtttagccgt      180
tttgacactg ttgtactact ggcagcggaa caccgcgatg acgtcagccc tacttctctc      240
tattatgatg tcaacgttca gggtaccgcg aatgtgctgg cggccatgga aaaaaatggc      300
gtaaaaaata tcatctttac cagttccgtt gctgtttatg gtttgaacaa acacaaccct      360
gacgaaaacc atccacacga cccittcaac cactacggca aaagtaagtg gcaggcagag      420
gaagtgcfcg gtgaatggta taacaaagca ccaacagaac gttcattaac catcatccgt      480
cetaccgtta tcttcgggtga acgaaccgcg ggtaacgtet ataacttgct gaaacagatc      540
gttggeggca agtttatgat ggtgggcgca gggactaact ataagtccat ggcttatggt      600
ggaaacattg ttgagtttat caagtacaaa ctgaagaatg ttgccgcagg ttatgaggtt      660
tataactatg ttgataagcc agacctgaac atgaaccagt tggttgctga agttgaacaa      720
agcctgaaca aaaagatccc ttctatgcac ttgccttacc cactaggaat gctgggtgga      780
tattgctttg atatcctgag caaaattacg ggcaaaaaat acgctgteag ctctgtgegc      840
gtgaaaaaat tctgcgcaac aacacagttt gacgcaacga aagtgcattc ttcaggtttt      900
gtggcaccgt atacgctgtc gcaaggtctg gatcgaaetc tgcagtatga attcgtccat      960
gccaaaaaag acgacataac gtttgtttct gag      993
```

<210> 8

<211> 331

<212> PRT

<213> 鲍氏志贺氏菌

<400> 8
 Met Asn Asp Asn Val Leu Leu Ile Gly Ala Ser Gly Phe Val Gly Thr
 1 5 10 15
 Arg Leu Leu Glu Thr Ala Ile Ala Asp Phe Asn Ile Lys Asn Leu Asp
 20 25 30
 Lys Gln Gln Ser His Phe Tyr Pro Ala Ile Thr Gln Ile Gly Asp Val
 35 40 45
 Arg Asp Gln Gln Ala Leu Asp Gln Ala Leu Ala Gly Phe Asp Thr Val
 50 55 60
 Val Leu Leu Ala Ala Glu His Arg Asp Asp Val Ser Pro Thr Ser Leu
 65 70 75 80
 Tyr Tyr Asp Val Asn Val Gln Gly Thr Arg Asn Val Leu Ala Ala Met
 85 90 95
 Glu Lys Asn Gly Val Lys Asn Ile Ile Phe Thr Ser Ser Val Ala Val
 100 105 110
 Tyr Gly Leu Asn Lys His Asn Pro Asp Glu Asn His Pro His Asp Pro
 115 120 125
 Phe Asn His Tyr Gly Lys Ser Lys Trp Gln Ala Glu Glu Val Leu Arg
 130 135 140
 [0007] Glu Trp Tyr Asn Lys Ala Pro Thr Glu Arg Ser Leu Thr Ile Ile Arg
 145 150 155 160
 Pro Thr Val Ile Phe Gly Glu Arg Asn Arg Gly Asn Val Tyr Asn Leu
 165 170 175
 Leu Lys Gln Ile Ala Gly Gly Lys Phe Met Met Val Gly Ala Gly Thr
 180 185 190
 Asn Tyr Lys Ser Met Ala Tyr Val Gly Asn Ile Val Glu Phe Ile Lys
 195 200 205
 Tyr Lys Leu Lys Asn Val Ala Ala Gly Tyr Glu Val Tyr Asn Tyr Val
 210 215 220
 Asp Lys Pro Asp Leu Asn Met Asn Gln Leu Val Ala Glu Val Glu Gln
 225 230 235 240
 Ser Leu Asn Lys Lys Ile Pro Ser Met His Leu Pro Tyr Pro Leu Gly
 245 250 255
 Met Leu Gly Gly Tyr Cys Phe Asp Ile Leu Ser Lys Ile Thr Gly Lys
 260 265 270
 Lys Tyr Ala Val Ser Ser Val Arg Val Lys Lys Phe Cys Ala Thr Thr
 275 280 285
 Gln Phe Asp Ala Thr Lys Val His Ser Ser Gly Phe Val Ala Pro Tyr
 290 295 300

Thr Leu Ser Gln Gly Leu Asp Arg Thr Leu Gln Tyr Glu Phe Val His
305 310 315 320

Ala Lys Lys Asp Asp Ile Thr Phe Val Ser Glu
325 330

<210> 9
<211> 993
<212> DNA
<213> 肠沙门氏菌

<400> 9
atgaacgata acgttttgct cattggtgct tccggattcg taggaacccg actecttgaa 60
acggcagttgg atgattttta tatcaagaac ctggataaac agcaaagcca tttctaccca 120
gagattacac acattggcga tgttcgtgac caacaaatcc ttgaccagac gttaggtgggt 180
tttgacaccg tagtactatt ggcctgggag catcgtgatg atgttagtcc tacctcgctt 240
tattatgatg tcaacgtcca gggaacgcgt aatgtactgg cggcgatgga aaaaaatggt 300
gtaaaaata tcatttttac cagttccgtt gcagtttatg gactcaacaa gaaaaatcct 360
gacgaaacgc accctcacga tccctttaat cttacggaa aaagtaaag gcaagcagaa 420
gaagtctcgc gtgagtgcca tgciaaagcg ccgaatgagc gttctttgac cataattcgt 480
cctaccgtta ttttcgggga gcgtaaccgc ggtaatgtat acaatctctt gaaacagatc 540
gttggtggta aatttgcgat ggttggtccg ggaactaact ataaatcaat ggettatgtt 600
ggtaatatcg ttgagtttat caaatcaaaa cteaagaatg ttacggcggg ctatgaagtt 660
[0008] tataattatg ttgataaacc tgaatgaat atgaatcaat tggttgctga agtagagcag 720
agcctgggca aaaaaatacc atcgatgcac ctccatatac cattaggtat gctgggggggt 780
tactgtttcg atatcctgag caaagtaacg ggcaagaagt acgctgtaag ttcggttcgt 840
gttaaaaaat tctgtgcgac aacgcagttt gatgcaacaa aagtgeattc ttctggtttt 900
gttgcgccat acacottatc tcagggttg gatcgtacac tgcaatatga attgttcat 960
gcaaagaaag atgacattac attcgtttca gag 993

<210> 10
<211> 331
<212> PRT
<213> 肠沙门氏菌

<400> 10
Met Asn Asp Asn Val Leu Leu Ile Gly Ala Ser Gly Phe Val Gly Thr
1 5 10 15

Arg Leu Leu Glu Thr Ala Val Asp Asp Phe Asn Ile Lys Asn Leu Asp
20 25 30

Lys Gln Gln Ser His Phe Tyr Pro Glu Ile Thr His Ile Gly Asp Val
35 40 45

Arg Asp Gln Gln Ile Leu Asp Gln Thr Leu Val Gly Phe Asp Thr Val
50 55 60

Val Leu Leu Ala Ala Glu His Arg Asp Asp Val Ser Pro Thr Ser Leu
65 70 75 80

	Tyr Tyr Asp Val Asn Val Gln Gly Thr Arg Asn Val Leu Ala Ala Met	
	85 90 95	
	Glu Lys Asn Gly Val Lys Asn Ile Ile Phe Thr Ser Ser Val Ala Val	
	100 105 110	
	Tyr Gly Leu Asn Lys Lys Asn Pro Asp Glu Thr His Pro His Asp Pro	
	115 120 125	
	Phe Asn His Tyr Gly Lys Ser Lys Trp Gln Ala Glu Glu Val Leu Arg	
	130 135 140	
	Glu Trp His Ala Lys Ala Pro Asn Glu Arg Ser Leu Thr Ile Ile Arg	
	145 150 155 160	
	Pro Thr Val Ile Phe Gly Glu Arg Asn Arg Gly Asn Val Tyr Asn Leu	
	165 170 175	
	Leu Lys Gln Ile Ala Gly Gly Lys Phe Ala Met Val Gly Pro Gly Thr	
	180 185 190	
	Asn Tyr Lys Ser Met Ala Tyr Val Gly Asn Ile Val Glu Phe Ile Lys	
	195 200 205	
	Phe Lys Leu Lys Asn Val Thr Ala Gly Tyr Glu Val Tyr Asn Tyr Val	
	210 215 220	
[0009]	Asp Lys Pro Asp Leu Asn Met Asn Gln Leu Val Ala Glu Val Glu Gln	
	225 230 235 240	
	Ser Leu Gly Lys Lys Ile Pro Ser Met His Leu Pro Tyr Pro Leu Gly	
	245 250 255	
	Met Leu Gly Gly Tyr Cys Phe Asp Ile Leu Ser Lys Val Thr Gly Lys	
	260 265 270	
	Lys Tyr Ala Val Ser Ser Val Arg Val Lys Lys Phe Cys Ala Thr Thr	
	275 280 285	
	Gln Phe Asp Ala Thr Lys Val His Ser Ser Gly Phe Val Ala Pro Tyr	
	290 295 300	
	Thr Leu Ser Gln Gly Leu Asp Arg Thr Leu Gln Tyr Glu Phe Val His	
	305 310 315 320	
	Ala Lys Lys Asp Asp Ile Thr Phe Val Ser Glu	
	325 330	
	<210> II	
	<211> 987	
	<212> DNA	
	<213> 空肠弯曲杆菌	
	<400> II	
	atgaaaattc ttattagcgg tgggtcaggt tatataggtt ctacatacttt aagacaattt	60
	ttaaaaacag atcatgaaat ttgtgtttta gataatcttt ctaagggttc taaaatcgca	120
	atagaagatt tgcaaaaaac aagagctttt aaatttttcg aacaagattt aagtgatttt	180

```

caaggcgtaa aagcattggt tgagagagaa aaatttgacg ctattgtgca ttttcagea    240
agcattgaag tttttgaaag tatgcaaaat cttttaaaat attatataaa caacactgtt    300
aatacgacaa atctcataga aacttggttg caaacggag igaataaatt tataatttct    360
tcaacggcgg ccacttatgg cgaaccacaa actcccggtg tgagcgaaac aagtccttta    420
gcacctatta atccttatgg gcgtagtaag cttatgagtg aagaagtttt gcgtgatgca    480
agtatggcaa atcctgaatt taagcattgt attttaagat attttaatgt tgcaggtgct    540
tgtatggatt ataactttagg acaacgetat ccaaaagcga ctttgcttat aaaagttgca    600
gttgaatgtg ccgcaggaaa acgtgataaa cttttcatat ttggcgatga ttatgataca    660
aaagatggta ctgcataag agattttatc catgtagatg atatttcaag tgcacattta    720
gcggctttgg attattttaa agagaatgaa agcaatgttt ttaatgtagg ttatggacat    780
ggttttagcg taaaagaagt gattgaagcg atgaaaaaag ttagcggagt ggattttaaa    840
gtagaacttg ccccaagcgg tgggggtgat cctagtgtat tgatttctga tgcaagtaaa    900
atcagaaatc ttacttcttg gcagcctaaa tatgatgatt tagagcttat ttgtaaactc    960
gcttttgatt gggaaaaaca gtgttaa    987

```

<210> 12

<211> 328

<212> PRT

<213> 空肠弯曲杆菌

<400> 12

```

Met Lys Ile Leu Ile Ser Gly Gly Ala Gly Tyr Ile Gly Ser His Thr
1           5           10          15

```

[0010]

```

Leu Arg Gln Phe Leu Lys Thr Asp His Glu Ile Cys Val Leu Asp Asn
20          25          30

```

```

Leu Ser Lys Gly Ser Lys Ile Ala Ile Glu Asp Leu Gln Lys Thr Arg
35          40          45

```

```

Ala Phe Lys Phe Phe Glu Gln Asp Leu Ser Asp Phe Gln Gly Val Lys
50          55          60

```

```

Ala Leu Phe Glu Arg Glu Lys Phe Asp Ala Ile Val His Phe Ala Ala
65          70          75          80

```

```

Ser Ile Glu Val Phe Glu Ser Met Gln Asn Pro Leu Lys Tyr Tyr Met
85          90          95

```

```

Asn Asn Thr Val Asn Thr Thr Asn Leu Ile Glu Thr Cys Leu Gln Thr
100         105         110

```

```

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

```

```

Pro Gln Thr Pro Val Val Ser Glu Thr Ser Pro Leu Ala Pro Ile Asn
130         135         140

```

```

Pro Tyr Gly Arg Ser Lys Leu Met Ser Glu Glu Val Leu Arg Asp Ala
145         150         155         160

```

```

Ser Met Ala Asn Pro Glu Phe Lys His Cys Ile Leu Arg Tyr Phe Asn

```

Val	Ala	Gly	Ala	Cys	Met	Asp	Tyr	Thr	Leu	Gly	Gln	Arg	Tyr	Pro	Lys	165	170	175
		180						185					190					
Ala	Thr	Leu	Leu	Ile	Lys	Val	Ala	Ala	Glu	Cys	Ala	Ala	Gly	Lys	Arg		200	205
	195																	
Asp	Lys	Leu	Phe	Ile	Phe	Gly	Asp	Asp	Tyr	Asp	Thr	Lys	Asp	Gly	Thr	210	215	220
Cys	Ile	Arg	Asp	Phe	Ile	His	Val	Asp	Asp	Ile	Ser	Ser	Ala	His	Leu	225	230	235
															240			
Ala	Ala	Leu	Asp	Tyr	Leu	Lys	Glu	Asn	Glu	Ser	Asn	Val	Phe	Asn	Val	245	250	255
Gly	Tyr	Gly	His	Gly	Phe	Ser	Val	Lys	Glu	Val	Ile	Glu	Ala	Met	Lys	260	265	270
Lys	Val	Ser	Gly	Val	Asp	Phe	Lys	Val	Glu	Leu	Ala	Pro	Arg	Arg	Ala	275	280	285
Gly	Asp	Pro	Ser	Val	Leu	Ile	Ser	Asp	Ala	Ser	Lys	Ile	Arg	Asn	Leu	290	295	300
Thr	Ser	Trp	Gln	Pro	Lys	Tyr	Asp	Asp	Leu	Glu	Leu	Ile	Cys	Lys	Ser	305	310	315
															320			
Ala	Phe	Asp	Trp	Glu	Lys	Gln	Cys									325		

[0011]

<210> 13
 <211> 1017
 <212> DNA
 <213> 大肠杆菌

<400> 13
 atgagagttc tggttaccgg tggtagcggg tacattggaa gtcataacctg tgtgcaatta 60
 ctgcaaaaacg gtcattgatg catcattctt gataacctct gtaacagtaa gcgcagcgta 120
 ctgcctgtta tegagcgttt aggcggcaaa catccaacgt ttgttgaagg cgatatctgt 180
 aacgaagcgt tgatgaccca gatcctgcac gatcacgcta tcgacaccgt gatecaette 240
 gccgggctga aagccgtggg cgaatcggtta caaaaaccgc tggaatatta cgacaacaat 300
 gtcaacggca ctctgcgcct gattagcgcc atgcgcgcgc ctaacgtcaa aaactttatt 360
 tttagctcct ccgccaccgt ttatggcgat cagcccaaaa ttccatacgt tgaaagcttc 420
 ccgaaccggca caccgcaaaag cccttacggc aaaagcaagc tgatgggtgga acagatccct 480
 accgatctgc aaaaagccca gccggactgg agcatigccc tgcctgcgta ctccaaccgc 540
 gttggcgcgc atccgtcggg cgatatgggc gaagatccgc aaggcaatcc gaataacctg 600
 atgccatata tgcgccaggt tgctgtaggg cgtcgcgact cgctggcgat ttttggtaac 660
 gattatccga ccgaagatgg tacttgcgta cgcgattaca tccacgtaat ggatctggcg 720
 gacggtaacg tcgtggcgat ggaanaactg gcgaacaagc caggcgtaaa catctacaac 780
 ctccggcgctg gcgtaggcaa cagcgtgcgt gacgtgggta atgccttcag caaagcctgc 840

ggcaaaccgg ttaattatca ttttgcaccg cgtcgcgagg ggcaccttcc ggcctacigg 900
 gcggacgccca gcaaagecga ccgtgaactg aactggcgcg taacgcgcac actcgatgaa 960
 atggcgcagg acacctggca ctggcagtcg cgccatccac agggatatcc cgattaa 1017

<210> 14
 <211> 338
 <212> PRT
 <213> 大肠杆菌

<400> 14
 Met Arg Val Leu Val Thr Gly Gly Ser Gly Tyr Ile Gly Ser His Thr
 1 5 10 15

Cys Val Gln Leu Leu Gln Asn Gly His Asp Val Ile Ile Leu Asp Asn
 20 25 30

Leu Cys Asn Ser Lys Arg Ser Val Leu Pro Val Ile Glu Arg Leu Gly
 35 40 45

Gly Lys His Pro Thr Phe Val Glu Gly Asp Ile Arg Asn Glu Ala Leu
 50 55 60

Met Thr Glu Ile Leu His Asp His Ala Ile Asp Thr Val Ile His Phe
 65 70 75 80

Ala Gly Leu Lys Ala Val Gly Glu Ser Val Gln Lys Pro Leu Glu Tyr
 85 90 95

[0012] Tyr Asp Asn Asn Val Asn Gly Thr Leu Arg Leu Ile Ser Ala Met Arg
 100 105 110

Ala Ala Asn Val Lys Asn Phe Ile Phe Ser Ser Ser Ala Thr Val Tyr
 115 120 125

Gly Asp Gln Pro Lys Ile Pro Tyr Val Glu Ser Phe Pro Thr Gly Thr
 130 135 140

Pro Gln Ser Pro Tyr Gly Lys Ser Lys Leu Met Val Glu Gln Ile Leu
 145 150 155 160

Thr Asp Leu Gln Lys Ala Gln Pro Asp Trp Ser Ile Ala Leu Leu Arg
 165 170 175

Tyr Phe Asn Pro Val Gly Ala His Pro Ser Gly Asp Met Gly Glu Asp
 180 185 190

Pro Gln Gly Ile Pro Asn Asn Leu Met Pro Tyr Ile Ala Gln Val Ala
 195 200 205

Val Gly Arg Arg Asp Ser Leu Ala Ile Phe Gly Asn Asp Tyr Pro Thr
 210 215 220

Glu Asp Gly Thr Gly Val Arg Asp Tyr Ile His Val Met Asp Leu Ala
 225 230 235 240

Asp Gly His Val Val Ala Met Glu Lys Leu Ala Asn Lys Pro Gly Val
 245 250 255

His Ile Tyr Asn Leu Gly Ala Gly Val Gly Asn Ser Val Leu Asp Val
260 265 270

Val Asn Ala Phe Ser Lys Ala Cys Gly Lys Pro Val Asn Tyr His Phe
275 280 285

Ala Pro Arg Arg Glu Gly Asp Leu Pro Ala Tyr Trp Ala Asp Ala Ser
290 295 300

Lys Ala Asp Arg Glu Leu Asn Trp Arg Val Thr Arg Thr Leu Asp Glu
305 310 315 320

Met Ala Gln Asp Thr Trp His Trp Gln Ser Arg His Pro Gln Gly Tyr
325 330 335

Pro Asp

<210> 15
<211> 1020
<212> DNA
<213> 大肠杆菌

[0013]

<400> 15
atggtgattt tcgtaacagg eggtcagga tatattggat cccataccat acttgagtta 60
cttaataatg gtcgatgtgt cgtttcgata gataattttg tcaattcctc tatagaatca 120
ttaaaaagag tagagcaaat aactaataag aaaattattt cttatcaagg tgatatecgt 180
gataaaaatc tacttgatga gattttttca agacaccata tcgatgetgt aattcacttt 240
gcatcgftaa aatcigtagg tgagtctaaag ttaaagccct tagagtatta ttctaataat 300
gttggtggaa ctttagtatt acttgaatgc atgaagagat ataacattaa taaaatgata 360
tttagctctt ctgctactgt ttaigggagt aacagtatcc ctccceatac ggaagataga 420
cgaattggtg aaactacaaa cccatatggg acatcgaaat ttataataga aataattttg 480
agtgattatt gigatagtga taataataaa tcagtaattg cactgcgtta ctttaatcca 540
atcgagacac ataagtcagg gatgattggt gaaaatecta acgggatccc taataatctg 600
gttccttata tatctaaagt tgcacaaaat caacttctg tattaatat ttatggcaac 660
gattatccaa ctaaagatgg tacaggagta agagactata tacatgtctg tgatttggt 720
aaagggcatg ttaaagcatt agaataatg tttttaaatg atgtcaattt tgaagctttt 780
aatttaggta ctggtcagg ttattctgtt ttagagattg taaaaatgtt tgagatagtc 840
actaaaaaga gtatacctgt tgcctattgt aatagacgtg agggggatgt tgcggagtca 900
tgggcgtctg ctgatttggt acataaaaag ctttctgga aagcggaaaa aaatttgaaa 960
gaaatgatcg aagatgtatg gcgttggcaa acaacaatc caaatggata taaaaataa 1020

<210> 16
<211> 339
<212> PRT
<213> 大肠杆菌

<400> 16
Met Val Ile Phe Val Thr Gly Gly Ala Gly Tyr Ile Gly Ser His Thr
1 5 10 15

Ile Leu Glu Leu Leu Asn Asn Gly His Asp Val Val Ser Ile Asp Asn
 20 25 30
 Phe Val Asn Ser Ser Ile Glu Ser Leu Lys Arg Val Glu Gln Ile Thr
 35 40 45
 Asn Lys Lys Ile Ile Ser Tyr Gln Gly Asp Ile Arg Asp Lys Asn Leu
 50 55 60
 Leu Asp Glu Ile Phe Ser Arg His His Ile Asp Ala Val Ile His Phe
 65 70 75 80
 Ala Ser Leu Lys Ser Val Gly Glu Ser Lys Leu Lys Pro Leu Glu Tyr
 85 90 95
 Tyr Ser Asn Asn Val Gly Gly Thr Leu Val Leu Leu Glu Cys Met Lys
 100 105 110
 Arg Tyr Asn Ile Asn Lys Met Ile Phe Ser Ser Ser Ala Thr Val Tyr
 115 120 125
 Gly Ser Asn Ser Ile Pro Pro His Thr Glu Asp Arg Arg Ile Gly Glu
 130 135 140
 Thr Thr Asn Pro Tyr Gly Thr Ser Lys Phe Ile Ile Glu Ile Ile Leu
 145 150 155 160
 [0014] Ser Asp Tyr Cys Asp Ser Asp Asn Asn Lys Ser Val Ile Ala Leu Arg
 165 170 175
 Tyr Phe Asn Pro Ile Gly Ala His Lys Ser Gly Met Ile Gly Glu Asn
 180 185 190
 Pro Asn Gly Ile Pro Asn Asn Leu Val Pro Tyr Ile Ser Lys Val Ala
 195 200 205
 Gln Asn Gln Leu Pro Val Leu Asn Ile Tyr Gly Asn Asp Tyr Pro Thr
 210 215 220
 Lys Asp Gly Thr Gly Val Arg Asp Tyr Ile His Val Cys Asp Leu Ala
 225 230 235 240
 Lys Gly His Val Lys Ala Leu Glu Tyr Met Phe Leu Asn Asp Val Asn
 245 250 255
 Tyr Glu Ala Phe Asn Leu Gly Thr Gly Gln Gly Tyr Ser Val Leu Glu
 260 265 270
 Ile Val Lys Met Phe Glu Ile Val Thr Lys Lys Ser Ile Pro Val Ala
 275 280 285
 Ile Cys Asn Arg Arg Glu Gly Asp Val Ala Glu Ser Trp Ala Ser Ala
 290 295 300
 Asp Leu Ala His Lys Lys Leu Ser Trp Lys Ala Glu Lys Asn Leu Lys
 305 310 315 320

Glu Met Ile Glu Asp Val Trp Arg Trp Gln Thr Asn Asn Pro Asn Gly
325 330 335

Tyr Lys Lys

<210> 17
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述: 合成
引物

<400> 17
aaacccggga tgaacgataa cgttttctgc 30

<210> 18
<211> 60
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述: 合成
引物

<400> 18
aaatctagat taagcgtaat ctggaacatc gtatgggtac tcagaaacaa acgttatgtc 60

[0015]

<210> 19
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述: 合成
引物

<400> 19
aaaccatgga tgaaaattct tattagcgg 29

<210> 20
<211> 57
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述: 合成
引物

<400> 20
aaatctagat taagcgtaat ctggaacatc gtaigggtag caetgttttt cccaatc 57

<210> 21
<211> 11
<212> DNA
<213> 人工序列

<220>
<223> 人工序列描述: 合成
寡核苷酸

<400> 21
aaaaagctag c 11

<210> 22
<211> 8

<212> DNA		
<213> 人工序列		
<220>		
<223> 人工序列描述: 合成寡核苷酸		
<400> 22		
ccgcgcgg		8
<210> 23		
<211> 7794		
<212> DNA		
<213> 人工序列		
<220>		
<223> 人工序列描述: 合成多核苷酸		
<400> 23		
tctacgggt ctgacgtca gtggaacgaa atcgatgagc tcgcacgaac ccagttgaca	60	
taagccigt ttggttcgtaa acigtaatgc aagtagcgta tgcgctcagc caactgggtc	120	
agaacctga ccgaacgcag cgggtggtaac ggccgagtg cggttttcat ggcttggtat	180	
gaetgtttt ttgtacagtc tagcctcggg catecaagct agctaagcgc gttacgcgt	240	
gggtcgatgt ttgatgttat ggaacagcaa cgatgttacg cagcagggtg gtgcacctaa	300	
aacaaagtta ggcagccgtt gtgctgtgtc ttctagtag ttgttgtggg gtaggcagtc	360	
agagctcgat ttgctgtcgc ccataataga ttcacaagaa ggattcgaca tgggtcaaag	420	
tagcgatgaa gccaacgctc ccgttcagg gcagtttgcc cttccctga gtgccacctt	480	
tggtttaggg gategcgtae gcaagaaatc tgggtccgct tggcagggtc aagtcgtcgg	540	
ttgttatgce aaaaactca ctctgaagg ctatgcggtc ggtccgaat cccaccagg	600	
ctcagtgcaa atttatectg ttgctgcaat tgaacgtgtg gcctaagcga tatcttagga	660	
tctcccatcg gtgatgtcgg cgatataggg gccagcaacc gcacctgtgg cgcgggtgat	720	
gcgggccacg atgcgtccgg cgtagaggat ctgctcatgt ttgacagctt atcctcgatg	780	
cataatgtgc ctgtcaaatg gacgaagcag ggattctgca aacctatgc tactccgtca	840	
agccgtcaat tgtctgattc gttaccaatt atgacaactt gacggctaca tcatcactt	900	
tttttcaca accggcacgg aactcgtcgc ggttggtccc ggtgcatttt ttaaataccc	960	
gcgagaaata gatttgatcg tcaaaaccaa cattgcgacc gacggtggcg ataggcatcc	1020	
gggtgtgtgt caaaagcagc ttgccttggc tgatacgttg gtcctcgcgc cagcttaaga	1080	
cgtaatccc taactgtcgg cggaaaagat gtgacagacg cgacggcgac aagcaaacat	1140	
gtgtgtcgac gctggcgata tcaaaattgc tgtctgccag gtgctcgtg atgtactgac	1200	
aagcctcgcg tacccgatta tccatcgggt gatggagcga ctctttaatc gcttccatgc	1260	
gcgcagtaaa caattgcata agcagattta tcgccagcag ctccgaatag cgccttccc	1320	
cttgcccggc gttaatgatt tgcccaaca ggctcgtgaa atgcggctgg tgcgttcat	1380	
ccgggcgaaa gaaccccgta ttggcaata ttgacggcca gtttaagccat tcatgccagt	1440	
aggcgcggcg acgaaagtaa acccaactgtt gataccattc gcgagcctcc ggatgacgac	1500	
cgtagtgtat aatctctcct ggccgggaaca gcaaaatata acccggtcgg caaacaatt	1560	
ctcgtccctg atttttccac accccctgac ccgaatggt gagattgaga atataacctt	1620	
tcatcccgag cggctcgtcg ataaaaaat ccagataacc gttggcctca atcggcgtta	1680	

[0016]

	aacccgccac cagatgggca ttaaacgagt atcccggcag caggggatca ttttgcgtt	1740
	cagccatact tttcatactc ccgccattca gagaagaaac caattgtcca tattgcatca	1800
	gacattgccg tcactgcgtc ttttactggc tttctctgct aaccaaaccg gtaacccgcg	1860
	ttattaaaag caitctgtaa caaagcggga ccaaagecat gacaaaaacg cgtaaacaaa	1920
	gtgtctataa tcacggcaga aaagtccaca ttgattatit gcacggcgtc acactttgct	1980
	atgccatage attttatcc ataagattag cggatcctac ctgacgcttt ttatcgcaac	2040
	tcctactgtt ttccatac ccgtttttt gggttagcag gaggaattca ccatggtacc	2100
	cgggatgaac gataacgttt tgcctatagg agcttcggga ttctaggaa cccgactact	2160
	tgaacggca attgtgact ttaatafcaa gaacctggac aaacagcaga gccacttta	2220
	tccagaaac acacagattg gcgatgttcg cgatcaacag gcactcgacc aggcgttagt	2280
	cggttttgac actgttgatc tactggcagc ggaacacgcg gatgacgtca gccctacttc	2340
	tcctatttat gatgtcaacg ttacgggtac ccgcaatgtg ctggcggcca tggaaaaaaa	2400
	tggcgtaaaa aatatcatct ttaccagttc cgttgcgttt tatggtttga acaaacacaa	2460
	ccctgacgaa aaccatccac acgaccttt caaccactac ggcaaaagta agtggcaggc	2520
	agaggaaagt ctgcgtgaat ggtataacaa agcaccaaca gaacgttcat taacctcat	2580
	ccgtcttacc gttatcttcg gtgaacgcaa ccgcggtaac gtctataact tgcgtaaaca	2640
	gacgcctggc ggcaagttaa tgatgggtgg cgcagggaact aactataagt ccatggetta	2700
	tgttggaac attgttgagt ttatcaagia caaactgaag aatgttgccg caggttatga	2760
	ggtttataac tactttgata agecagacct gaacatgaac cagttggttg ctgaagtiga	2820
[0017]	acaaagccgt aacaaaaaga tcccttctat gcaettgect taccacttag gaatgcggg	2880
	tggatattgc ttigataacc tgagcaaaat tacgggcaaa aaatacgtg tcagctcagt	2940
	gcgcgtgaaa aaattctgcg caacaacaca gtttgacgca acgaaagtgc attcttcagg	3000
	ttttgtggca ccgtatacgc tgcgcgaagg tcctggatga acaactcagt atgaattcgt	3060
	tcatgcaaaa aaagacgaca taacgtttgt ttctgagtac ccatacgtg ttccagatta	3120
	cgtttaatct agagtcgacc tgcaggcatg caagettgce tgttttggeg gatgagagaa	3180
	gattttcagc ctgatacaga ttaaateaga acgcagaage ggtctgataa aacagaattt	3240
	gcctggcgcc agtagcgcg tggctcccac tgaccccatg ccgaactcag aagtgaacg	3300
	ccgtagcgcc gatggtagtg tgggtctccc ccatgcgaga gtagggaact gccaggcatc	3360
	aaataaaacg aaaggctcag tcgaaagact gggectttcg ttttatctgt tgtttgtcgg	3420
	tgaacgtctt cctgagtagg acaaatccgc cgggagcgga ttgaaagtt gcaagcaac	3480
	ggcccgagg gtggcgaggc ggacgccegc cataaactgc caggcatcaa attaacgaga	3540
	aggccatcct gacggatggc ctttttgcgt ttctacaaac tcttccactc actacagcag	3600
	agccatttaa acaacatccc cteccccctt ccaecgcgtc agacgccegt agcagccgc	3660
	tacgggcttt ttcatgccc gccctagcgt ccaagcctca cggcgcgct cggectctct	3720
	ggcgcccttc tgccgtgag gtctgcctcg tgaagaaggt gttgtgact cataccaggc	3780
	ctgaatgcgc ccatcatcca gccagaaagt gagggagcca cggttgatga gagctttgtt	3840
	gtaggiggac cagltgggtga ttttgaactt ttgctttgcc acggaacggt ctgcgttgtc	3900
	gggaagatgc gtgatctgat cttcaacac agcaaaagtt cgatttatcc aacaaagccg	3960

	ccgtcccgtc aagtcagcgt aatgcctcgc cagtggtaca accaattaac caattctgat	4020
	tagaaaaact cctcgagcat caaatgaac tgcaatttat tcataacagg attatcaata	4080
	ccataTTTTT gaaaaagccg tttctgtaat gaaggagaaa actcaccgag gcagttccat	4140
	aggatggcaa gatcctggta tcggctcgcg attccgactc gtccaacatc aatacaacct	4200
	attaatttcc cctcgtaaaa aataaggita tcaagcgaga aatcaccatg agtgacgact	4260
	gaatccgggtg agaattggca aagctaaaaa ggccgtaafa tccagctgaa cggctcgggt	4320
	ataggatcat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga	4380
	tatatcaacg gtggtatata cagtgatttt tttctccatt ttagcttccct tagctcctga	4440
	aaatctcgat aactcaaaaa atacgcccg tagtgatctt atttcattat ggtgaaagtt	4500
	ggaacctctt acgtgccgat caacgtctca ttttcgcaa aagttggccc agggcttccc	4560
	ggtatcaaca gggacaccag gatttatita ttctcggaag tgatcttccg tcacaggtat	4620
	ttatctgaag acgaaagggc ctctgtatac gctatttttt ataggtaat gtcagtataa	4680
	taatggttct ttagacgtca ggtggcactt ttccgggaaa tgtgcgcgcc cgcgttccctg	4740
	ctggcgctgg gctgttttct ggccgtggac ttcccgctgt tccgtcagca gcttttcgcc	4800
	cacggccttg atgatcgcg cggtccttggc ctgcataacc cgattcaacg gcccagggc	4860
	gtccagaacg ggttcaggc gctcccgaa gtctcgggcc gtctcttggg ctgtagcgcc	4920
	cttcttgcgc atctcacgcg ctctcgggc ggctgttagg gcaggctcat acccctgcgc	4980
	aaccgctttt gtcagccgtt cggccacggc ttccggcgtc tcaacgcgt ttgagattcc	5040
[0018]	cagcttttgc gccaatccct gcggtgcata ggccgctggc tcgaccgctt gcgggctgat	5100
	ggtgacgtgg ccaactgggt gccgcaccag ggctctgtag aacgcctgaa tgcgcgtgtg	5160
	acgtgccttg ctgcctcga tgcctcggtg cagccctaga tcggccacag cggccgcaaa	5220
	cgtggctcgg tcgcggtgca tctgcgtttt gttgccgatg aactccttgg ccgacagcct	5280
	ggcgtctcgc gtcagcgcca ccaagaaagg ggtcatgtgc gggtcgggtt cgtcacggtg	5340
	gatgtcggcc gtcacgatgc gatccgcccc gtacttgtec gccagccact tgtgcgctt	5400
	ctogaagaac gcgcctgct gttcttggct ggccgacttc caccattccg ggtcggcgt	5460
	catgacgtac tcgaccgcca acacagcgtc ctgcgcgcg ttctctggca gcaactcgcg	5520
	cagtcggccc atcgttccat cggctgtgct ggccgcccag tgcctgttct ctggcgtcct	5580
	gttggcgta cgtttggggc tctgcgcgc cgggtaggcg tgcctgagac tggccgcac	5640
	gttgccatt ttgcagct tcttgatcg catgatcgcg tatgcgcca tgcctgcgcc	5700
	tccttttttg tgccaacgc gctcgacggg ggcagcgcaa ggccgtgcct ccggcgggcc	5760
	actcaatgct tgagtatact cactagaact tgcctcgcaa agtcgtgacc gccacggcg	5820
	gttgcggcgc cctacgggtt tgcctccgg gcttcgccc gcgcggctgc tgcgtccct	5880
	tgcagcccc tggatatgtg gacgatggcc gcgagcgccc accgcttggc tgccttgcct	5940
	cggcccggtg acaacctgc tggacaagct gatggacagg ctgcgcctgc ccacagctt	6000
	gaccacaggg attgcccacc ggtacccag ccttcgacca catacccacc ggtcccaact	6060
	gcgcggcctg cggccttgc ccaatcaatt ttttaatttt ctctggggaa aagcctccgg	6120
	cctgcggcct gcgcgttcg ctgcccgtt ggacaccaag tggaggcggt gtaaggctc	6180
	gcgcagcgac cgcgcagcg ctggtccttg acgcgcctgg aacgacccaa gccatgcga	6240

	gtgggggcag tgaaggcga agcecgcccg cctgcccccc gagcctcagc gcggcgagtg	6300
	cgggggttcc aagggggcag egccaccttg ggcaaggccg aaggccgcgc agtcgateaa	6360
	caagccccgg agggggccact ttttgccgga gggggagccg cgcgaaggc gtgggggaac	6420
	ccgcaggggg tgcccttcct tgggcaccaa agaactagat atagggcgaa atgcgaaaga	6480
	cttaaaaaatc aacaacttaa aaaagggggg tacgcaacag ctcatctggg caccceccgc	6540
	aatagctcat tgcgtaggtt aaagaaaatc tgtaattgac tgccactttt acgcaacgca	6600
	taattgttgt cgcgcigccg aaaagtigca gctgattgeg catggtgccg caaccgtgcg	6660
	geacctacc gcatggagat aagcatggcc acgcagtcca gagaaatcgg cattcaagcc	6720
	aagaacaagc ccggtcactg ggtgcaaacg gaacgcaaag cgcattgaggc gtgggcccgg	6780
	cttattgcga ggaaaccac ggccgcaatg ctgctgcac accctcgtgc gcagatgggc	6840
	caccagaacg ccgtggtgtt cagccagaag acactttcca agctcctcgg acgttctttg	6900
	cggacggtcc aatacgcagt caaggacttg gtggccgagc gctggatctc cgctcgtgaag	6960
	ctcaacggcc ccggcacctg gtcggcctac gtggtcaatg acccgtggc gtggggccag	7020
	ccccgcgacc agttgcgctt gtcggtgttc agtgcccgcc tgggtgttga tcacgacgac	7080
	caggacgaat cgcgtgtggg gcatggcgac ctgcgcgca tcccgacct gtatccgggc	7140
	gagcagcaac taccgaccgg ccccgccgag gagccgcca gccagcccg cattccgggc	7200
	atggaaccag acctgccagc cttagccgaa acggaggaat gggaacggcg cgggcagcag	7260
	cgcctgcga tgcccgatga gccgtgtttt ctggacgatg gcgagccgtt ggagccgccc	7320
[0019]	acacgggtca cgcctgcgcg ccggtiagcact ttgggttgcg cagcaaccg taagtgcgt	7380
	gttccagact atcggtgtga gccgcctcgc cgcctatac cttgtctgcc tcccgcgtt	7440
	gcgtcgggt gcatggagcc gggccacctc gacctgaatg gaagccggcg gcacctcgt	7500
	aacggattca ccgtttttat caggtcttgg gaggcagaat aaatgatcat atcgtcaatt	7560
	attacctcca cggggagagc ctgagcaaac tggcctcagg catttgagaa gcacaaggtc	7620
	acactgttc cggtagtcaa taaaccgta aaccagcaat agacataagc ggctatttaa	7680
	cgacctgcc ctgaaccgac gaccgggtcg aatttcttt cgaatttctg ccattcatcc	7740
	gcttattate acttattcag gcgtagcacc aggcgtttaa gtcgaccaat aacc	7794
	<210> 24	
	<211> 7776	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 人工序列描述: 合成 多核苷酸	
	<400> 24	
	tctacgggt ctgacgtca gtggaacgaa atcgatgagc tcgcacgaac ccagttgaca	60
	taagcctgtt cgttcgtaa actgtaatgc aagtagcgta tgcgtcagc caactggtcc	120
	agaaccttga ccgaacgcag cgggtgtaac ggccgagtg cggttttcat ggcttgttat	180
	gactgttttt ttgtacagtc tagcctcggg catccaagct agctaagcgc gttacgcgt	240
	gggtcgatgt ttgatgttat ggaacagcaa cgatgttaag cagcagggta gtcgcctaa	300
	aacaaagtta ggcagccgtt gtcgtggtgc ttctagtag ttgtgtggg gtaggcagtc	360

	agagctcgat ttgcttgctg ccataataga ttacacaaga ggattcgaca tgggtcaaaag	420
	tagcgatgaa gccaaacgctc ccgttgccagg gcagtttgcg ctccccctga gtgccacett	480
	tggcttaggg gatcgcgtae gcaagaaatc tgggtccgct tggcagggtc aagtcgtcgg	540
	ttggfattgc acaaaactca ctccctgaagg ctatgcgggc gagtccgaat cccaccagg	600
	ctcagtgcaa atttatcctg ttgctgcaact tgaacgtgtg gcctaagcga tatcttagga	660
	tctcccatcg gtgatgtcgg cgatatagge gccagcaacc gcaactgtgg cgcgggtgat	720
	gcgggccacg atgcgtccgg cgtagaggat ctgctcatgt ttgacagett atcatcgatg	780
	cataatgtgc ctgtcaaatg gacgaagcag ggattctgca aaccttatgc tactccgtca	840
	agccgtcaat tgtctgattc gttaccaatt atgacaactt gacggetaca tcaticactt	900
	tttcttcaca accggcacgg aactcgcctg ggctggcccc ggtgcatttt ttaatacccc	960
	gggagaaata gatttgatcg tcaaaaccaa cattgcgacc gacgggtggcg ataggcatcc	1020
	gggtgggtgt caaaagcage ttegcctggc tgatacgttg gtectcgcgc cagettaaga	1080
	cgctaatacc taactgtcgg cggaaaagat gtgacagacg cgacggcgac aagcaaacat	1140
	gctgtgcgac gctggcgata tcaaaattgc tgtctgccag gtgatectg atgtactgac	1200
	aagcctcgcg tacccgatga tccatcggtg gatggagcga ctctttaac gtctccatgc	1260
	gcgcagtaa caattgtca agcagattta tgcgcagcag ctccgaatag cgccttccc	1320
	cttgcggcgc gttaatgatt tgcctcaaca ggctcgtgaa atgcggctgg tgcgttcat	1380
	ccggcgcaaa gaaccccgta ttggcaaaata ttgacggcca gtttaagccat tcatgccagt	1440
[0020]	aggcgcgcgg acgaaagtaa acccactggg gataccattc gcgagcctcc ggatgacgac	1500
	cgtagtgtat aatctctcct ggcggaaca gcaaaatate acccggtcgg caaacaatt	1560
	ctctgccctg atttttcacc acccctgac cgcgaatggt gagattgaga atataacctt	1620
	tcattcccag cggtcggctg ataaaaaat cgagataacc gttagcctca atcgcggtta	1680
	aaaccgccac cagatgggca ttaaacgagt atcccggcag caggggatea ttttgcgtt	1740
	cagccatact ttccatactc ccgccattca gagaagaac caattgtcca tattgcatca	1800
	gacattgcg tcactgcgtc ttttactggt tcttctcgtt aaccaaacgg gtaacccgc	1860
	ttattaaaag cattctgtaa caaagcggga ccaaagccat gacaaaaacg cgtaaacaaa	1920
	gtgtctataa tcacggcaga aaagtcaca ttgattattt gcacggcgtc acactttgct	1980
	atgccatage atttttatcc ataagattag cggatcctac ctgacgcttt ttatcgcaac	2040
	tctctactgt ttctccatac ccgttttttt gggttagcag gaggaattca ccatggatga	2100
	aaattcttat tagcgggtgt gcaggttata taggttctca tactttaaga caatttttaa	2160
	aaacagatca tgaattttgt gttttagata atctttctaa gggttctaaa atcgcaatag	2220
	aagatttgca aaaaataaga acttttaaat tttttgaaca agatttaagt gattttcaag	2280
	gcgtaaaage attgtttgag agagaaaaat ttgacgctat tgtgcatttt gcacgcagca	2340
	ttgaagtgtt tgaagtgatg caaaacctt taaagtatta tatgaataac actgttaata	2400
	cgacaaatct catcgaaact tgtttgcaaa ctggagtga taaatttata ttttcttcaa	2460
	cggcagccac ttatggcgaa ccacaaactc ccgttgtgag cgaacaaagt cctttagcac	2520
	ctattaatcc ttatggcggt agtaagctta tgagcgaaga ggttttgcgt gatgcaagia	2580
	tggcaaatcc tgaatttaag cattgtattt taagatattt taatgttgca ggtgcttgca	2640

[0021]

tggattatac ttiaggacaa cgctatccaa aagegacttt gcttataaaa gttgcagctg	2700
aatgtgccgc agaaaaaegt aataaacctt tcataatttg egatgattat gatacaaaag	2760
atggcaacttg cataagagat tttatccatg tggagatata ttaagtgcg catttategg	2820
ctttggatta tttaaaagag aatgaaagca atgtttttta ttaggttat ggacatggtt	2880
ttagcgtaaa agaagtgatt gaagcgatga aaaaagttag cggagtggat tttaaagtag	2940
aaactgcccc acgccgtgcg ggtgatecct gtgtattgat ttctgatgca agtaaaatca	3000
gaaatcttac ttcttggcag cctaaataig atgatttagg gcttatttgt aaatctgett	3060
ttgattggga aaaacagtgc taccatacag atgttcaga ttacgcttaa tctagagtcg	3120
acctgcagcg atgcaagctt ggctgitttg gcggatgaga gaagatttcc agccatgatac	3180
agattaaatc agaacgcaga agcggctcga taaaacagaa ttgacctggc ggcagtagcg	3240
cgggtggtccc acctgacccc atgccgaact cagaagtga acgccgtagc gccgatggta	3300
gtgtggggtc tccccatgag agagtaggga actgccagcg atcaataaaa acgaaaggct	3360
cagtcgaaag actgggacct tegttttatc tgttgttgt cgtgaacgc tctctgagt	3420
aggacaaatc cgcggggagc ggatttgaac gttgcgaagc aacggcccg agggtggcg	3480
gcaggacgcc cgccataaac tgcaggcat caaattaagc agaaggccat cctgaaggat	3540
ggcctttttg cgtttctaca aactcttcca ctactacag cagagccatt taaacaacat	3600
ccccccccc ttccaacgc gtcagacgcc cgtacagcc cgtacgggc ttttcatgc	3660
cctgccctag cgtccaagcc tcacggccgc gctcgccctc tctggcgcc ttctggcgt	3720
gaggtctgcc tctgaagaa ggtgttgcgt actcatacca ggctgaatc gcccacatc	3780
ccagccagaa agtgaggag ccacgggtga tgagagcttt gttgtaggtg gaccagttag	3840
tgattttgaa cttttgctt gccacggaac ggtctgcgtt gtcgggaaga tgcgtgatc	3900
gatecttcaa ctacgaaaa gtctgattta ttcaacaaag ccgcctccc gtcagtcag	3960
cgtaatgctc tgcagtggtt acaaccaatt aaccaattct gattagaaaa actcatcgag	4020
catcaaatga aactgcaatt tattcatatc aggtattatc ataccatatt ttgaaaaag	4080
ccgtttctgtaatgaaggag aaaactcacc gaggcagttc cataggatgg caagatcctg	4140
gtatcggtct gegattccga ctctccaac atcaatacaa cctattaatt tccccctgc	4200
aaaaataagg ttatcaagcg agaaatcacc atgagtgcg actgaatccg gtgagaatgg	4260
caaaagctaa aaaggccgta ataccagct gaacggctcgt gttataggta cattgagcaa	4320
ctgactgaaa tgcctcaaaa tgttctttac gatccattg ggatatatca acggtggtat	4380
atccagtgat tttttctcc attttagctt ccttagctcc tgaanaatctc gataactcaa	4440
aaaatacgc cggtagtgat cttatttcatt tatggigaaa gttggaacct cttacgtgcc	4500
gatcaacgctc tcattttccg caaaagttgg cccagggtt cccggtatca acaggacac	4560
caggatttat ttattctgcg aagtgatctt ccgtcacagg tatttatctg aagacgaaag	4620
ggcctcgiga tacgctatt tttatagggt aatgtcatga taataatggt ttcttagaag	4680
tcaggtggca ctttccgggg aaatlgcgc gcccgcttc ctgctggcgc tggcctgtt	4740
tctggcgtg gacttccgc tgttccgta gcagcttttc gccacggcc ttgatgatcg	4800
cggcggtctt ggctgcata tccgattca acggccccag ggcgtccaga acgggttca	4860
ggcgctccg aaggtctcgg gccgtctctt gggttgatc ggcctcttg cgcactcac	4920

	gcgtctctgc ggccggcctgt agggcaggct catacccttg ccgaaccgct ttigtacgcc	4980
	ggtcggccac ggtctccgga gctcaacgc gctttgagat tcccagcttt tggccaatc	5040
	cctgcgggtgc ataggcgctt ggctcgaccg ctigcgggct gatggtgacg tggcccactg	5100
	gtggccgcgc cagggcctcg tagaacgctt gaatgcgcgt gtgacgtgcc ttgtgcctt	5160
	cgatccccg ttgcagccct agatcgcca cagcggccgc aaacgtggtc tggtcgccc	5220
	tcacttgcgc ttgtttgcgc atgaactcct tggccgacag cctgcgcgcc tggtcagcg	5280
	gcaccacgaa cgcggctcat tggggctgg ttctgtacg gtggatgctg gccgtacga	5340
	tgcgacccg cccgtacttg tccgccagcc acctgtgcgc ctctcgaag aacccgcct	5400
	gctgtttctg gctggccgac ttccaccatt ccgggtggc cgtcatgacg tactcgaccg	5460
	ccaacacagc gtccttgcgc cgtttctcgt gcagcaactc gcgcagtcgg cccatgcctt	5520
	catcggtgtg gctggccgcc cagtgctcgt tctctggcgt cctgtggcg tcagcgttgg	5580
	gggtctcgcg ctgcggtag gctgtcttga gactggccgc cagcttgccc attttgccca	5640
	gcttcttgcg tcgcatgate gcttatgccg ccattgcctgc cctcccttt tgggtccaa	5700
	ccggctcgac gggggcagcg caaggcgtg cctccggcgg gccactcaat gcttgagtat	5760
	actactaga ctttgcctcg caaagtcgtg accgcctacg gcggctggcg cgcctacgg	5820
	gcttgcctc cgggcttcgc cctgcgggtt cgtgcgcgc ccttgccagc cgttgatat	5880
	gtggacgatg gccgcgagcg gccaccgct ggtcgcctt cctcgcccg tggacaacc	5940
	tgttgacaa gctgatggac aggtgcgcgc tggccagcg cttgaccaca gggattgcc	6000
[0022]	accggctacc cagcttcca ccacataccc accgctcca actgcgcgc ctcggcctt	6060
	gccccatcaa ttttttaat ttctctggg gaaaagcct cggcctgcgg cctgcgcgt	6120
	tcgtttgcgc gttagacacc aagtggaaag cgggtcaagg ctgcgcgcgc gaccgcgcg	6180
	cggcttgccc ttgacgcgcc tggaaagacc caagcctatg cgagtggggg cagtcgaag	6240
	cgaagccgc cgcctgcgc cccgagctc accgcccga gtgcgggggt tccaggggg	6300
	cagcccccac ttgggcaagg ccgaaggccg cgcagtcgat caacaagccc cggaggggc	6360
	actttttgac ggagggggag ccgcgcgaa ggcgtggggg aacccgcag ggtgcccctt	6420
	ctttgggcac caaagaacta gatataggg gaaatgcga agacttaaaa atcaacaact	6480
	taaaaaagg ggttacgaa cagtcattg cggcaccccc cgaatagct cattgctag	6540
	gttaaagaaa atctgtaatt gactgcact tttaacgaa gcataattgt tgcgcgcgt	6600
	ccgaaagtt gcagctgatt ggcgatgtg ccgcaaccgt gcggcaccct accgatgga	6660
	gataagcatg gccacgcagt ccagagaaat cggcattcaa gccagaaca agcccggtca	6720
	ctgggtgcaa accgaacgca aagcgcata ggcgtgggccc gggcttattg cgaggaaacc	6780
	cacggcgcca atgtgtgc atcacctcgt ggcgcagatg gccaccaga acccgtggt	6840
	ggtcagccag aagacaattt ccaagtcct cggacgttct ttgcggacgg tccaaatcgc	6900
	agtaaggac ttgttgccc agcgttgat ctccgtcgtg aagctcaag gccccggcac	6960
	cgtgtcggcc tacgtgttca atgaccgct ggcgtggggc cagcccccgc accagttgcg	7020
	cctgtcgggt ttacgtgcc ccgtgtgtgt tgatcacgac gaccaggacg aatcgtgtt	7080
	ggggcatggc gacctgcgc gcateccgac cctgtatccg ggcagacgc aactaccgac	7140
	cggcccccgc gaggagccgc ccagccagcc cggcattccg gcatggaac cagacctgcc	7200

```

agccttgacc gaaacggagg aatgggaacg gcgcgggcag cagcgcctgc cgatgcccca 7260
tgagccgtgt tttctggacg atggcgagcc gttggagccg ccgacacggg tcacgtgccc 7320
gcgcggtagt cacttgggtt gcgcagcaac ccgtaagtgc gctgttcag actatcggtt 7380
gtagecgcct cgcgcgccta taccttgtct gcctccccgc gttggtgcgc ggtgcatgga 7440
gcgcggccac ctgcacctga atggaagccg gcggcaacct gctaacggat tcacctttt 7500
tatacggctc tgggagcgag aataaatgat catatcgta attattacct ccacggggag 7560
agcctgagca aactggcctc aggcatttga gaagcacacg gtcacactgc ttccggtagt 7620
caataaacgg gtaaacccagc aatagacata agcggctatt taacgacctt gccctgaacc 7680
gacgacgggg tcgaatttgc ttctgaattt ctgccattca tccgcttatt atcacttatt 7740
caggcgtagc accaggcgtt taagtcgacc aataac 7776

```

<210> 25

<211> 643

<212> PRT

<213> 人工序列

<220>

<223> 人工序列描述: 合成多肽

<400> 25

```

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
1          5          10          15

```

```

Ala Ser Ala Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys
          20          25          30

```

[0023]

```

Ala Cys Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser
          35          40          45

```

```

Val Asp Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr
          50          55          60

```

```

Ser Met Val Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp
          65          70          75          80

```

```

Asn Ala Leu Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly
          85          90          95

```

```

Gly Val Glu Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala
          100          105          110

```

```

Arg Gly Ser Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys
          115          120          125

```

```

Pro Ser Asn Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln
          130          135          140

```

```

Leu Ser His Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu
          145          150          155          160

```

```

Leu Ala Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu
          165          170          175

```

```

Ser Asn Glu Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser

```

	180	185	190
	Val Val Met Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu 195 200 205		
	Trp Ala Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val 210 215 220		
	Tyr Asn Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu 225 230 235 240		
	Gly Lys Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu 245 250 255		
	Asp Ile Lys Asp Asn Asn Asn Ser Thr Pro Thr Val Ile Ser His Arg 260 265 270		
	Leu His Phe Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln 275 280 285		
	Ala Cys His Leu Pro Leu Glu Ala Phe Thr Arg His Arg Gln Pro Arg 290 295 300		
	Gly Trp Glu Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val 305 310 315 320		
	Ala Leu Tyr Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val 325 330 335		
[0024]	Ile Arg Asn Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu 340 345 350		
	Ala Ile Arg Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala 355 360 365		
	Ala Ala Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu 370 375 380		
	Ala Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala 385 390 395 400		
	Lys Asp Gln Asn Arg Thr Lys Gly Glu Cys Ala Gly Pro Ala Asp Ser 405 410 415		
	Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu 420 425 430		
	Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp 435 440 445		
	Thr Val Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly 450 455 460		
	Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser 465 470 475 480		
	Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile		

	485	490	495
	Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr 500 505 510		
	Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala 515 520 525		
	Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg 530 535 540		
	Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu Arg 545 550 555 560		
	Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro 565 570 575		
	Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala Glu 580 585 590		
	Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val 595 600 605		
	Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile 610 615 620		
[0025]	Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu 625 630 635 640		
	Asp Leu Lys		
	<210> 26 <211> 722 <212> PRT <213> 空肠弯曲杆菌		
	<400> 26 Met Leu Lys Lys Glu Tyr Leu Lys Asn Pro Tyr Leu Val Leu Phe Ala 1 5 10 15		
	Met Ile Ile Leu Ala Tyr Val Phe Ser Val Phe Cys Arg Phe Tyr Trp 20 25 30		
	Val Trp Trp Ala Ser Glu Phe Asn Glu Tyr Phe Phe Asn Asn Gln Leu 35 40 45		
	Met Ile Ile Ser Asn Asp Gly Tyr Ala Phe Ala Glu Gly Ala Arg Asp 50 55 60		
	Met Ile Ala Gly Phe His Gln Pro Asn Asp Leu Ser Tyr Tyr Gly Ser 65 70 75 80		
	Ser Leu Ser Ala Leu Thr Tyr Trp Leu Tyr Lys Ile Thr Pro Phe Ser 85 90 95		
	Phe Glu Ser Ile Ile Leu Tyr Met Ser Thr Phe Leu Ser Ser Leu Val 100 105 110		

Val Ile Pro Thr Ile Leu Leu Ala Asn Glu Tyr Lys Arg Pro Leu Met
 115 120 125
 Gly Phe Val Ala Ala Leu Leu Ala Ser Ile Ala Asn Ser Tyr Tyr Asn
 130 135 140
 Arg Thr Met Ser Gly Tyr Tyr Asp Thr Asp Met Leu Val Ile Val Leu
 145 150 155 160
 Pro Met Phe Ile Leu Phe Phe Met Val Arg Met Ile Leu Lys Lys Asp
 165 170 175
 Phe Phe Ser Leu Ile Ala Leu Pro Leu Phe Ile Gly Ile Tyr Leu Trp
 180 185 190
 Trp Tyr Pro Ser Ser Tyr Thr Leu Asn Val Ala Leu Ile Gly Leu Phe
 195 200 205
 Leu Ile Tyr Thr Leu Ile Phe His Arg Lys Glu Lys Ile Phe Tyr Ile
 210 215 220
 Ala Val Ile Leu Ser Ser Leu Thr Leu Ser Asn Ile Ala Trp Phe Tyr
 225 230 235 240
 Gln Ser Ala Ile Ile Val Ile Leu Phe Ala Leu Phe Ala Leu Glu Gln
 245 250 255
 [0026] Lys Arg Leu Asn Phe Met Ile Ile Gly Ile Leu Gly Ser Ala Thr Leu
 260 265 270
 Ile Phe Leu Ile Leu Ser Gly Gly Val Asp Pro Ile Leu Tyr Gln Leu
 275 280 285
 Lys Phe Tyr Ile Phe Arg Ser Asp Glu Ser Ala Asn Leu Thr Gln Gly
 290 295 300
 Phe Met Tyr Phe Asn Val Asn Gln Thr Ile Gln Glu Val Glu Asn Val
 305 310 315 320
 Asp Leu Ser Glu Phe Met Arg Arg Ile Ser Gly Ser Gln Ile Val Phe
 325 330 335
 Leu Phe Ser Leu Phe Gly Phe Val Trp Leu Leu Arg Lys His Lys Ser
 340 345 350
 Met Ile Met Ala Leu Pro Ile Leu Val Leu Gly Phe Leu Ala Leu Lys
 355 360 365
 Gly Gly Leu Arg Phe Thr Ile Tyr Ser Val Pro Val Met Ala Leu Gly
 370 375 380
 Phe Gly Phe Leu Leu Ser Glu Phe Lys Ala Ile Met Val Lys Lys Tyr
 385 390 395 400
 Ser Gln Leu Thr Ser Asn Val Cys Ile Val Phe Ala Thr Ile Leu Thr
 405 410 415

	Leu	Ala	Pro	Val	Phe	Ile	His	Ile	Tyr	Asn	Tyr	Lys	Ala	Pro	Thr	Val	
				420					425					430			
	Phe	Ser	Gln	Asn	Glu	Ala	Ser	Leu	Leu	Asn	Gln	Leu	Lys	Asn	Ile	Ala	
			435					440					445				
	Asn	Arg	Glu	Asp	Tyr	Val	Val	Thr	Trp	Ala	Ala	Tyr	Gly	Tyr	Pro	Val	
			450				455					460					
	Arg	Tyr	Tyr	Ser	Asp	Val	Lys	Thr	Leu	Val	Asp	Gly	Gly	Lys	His	Leu	
	465					470					475					480	
	Gly	Lys	Asp	Asn	Phe	Phe	Pro	Ser	Phe	Ala	Leu	Ser	Lys	Asp	Glu	Gln	
					485					490					495		
	Ala	Ala	Ala	Asn	Met	Ala	Arg	Leu	Ser	Val	Glu	Tyr	Thr	Glu	Lys	Ser	
				500				505						510			
	Phe	Tyr	Ala	Pro	Gln	Asn	Asp	Ile	Leu	Lys	Thr	Asp	Ile	Leu	Gln	Ala	
			515				520					525					
	Met	Met	Lys	Asp	Tyr	Asn	Gln	Ser	Asn	Val	Asp	Leu	Phe	Leu	Ala	Ser	
		530					535					540					
	Leu	Ser	Lys	Pro	Asp	Phe	Lys	Ile	Asp	Thr	Pro	Lys	Thr	Arg	Asp	Ile	
	545					550				555					560		
[0027]	Tyr	Leu	Tyr	Met	Pro	Ala	Arg	Met	Ser	Leu	Ile	Phe	Ser	Thr	Val	Ala	
				565						570					575		
	Ser	Phe	Ser	Phe	Ile	Asn	Leu	Asp	Thr	Gly	Val	Leu	Asp	Lys	Pro	Phe	
				580				585						590			
	Thr	Phe	Ser	Thr	Ala	Tyr	Pro	Leu	Asp	Val	Lys	Asn	Gly	Glu	Ile	Tyr	
			595				600						605				
	Leu	Ser	Asn	Gly	Val	Val	Leu	Ser	Asp	Asp	Phe	Arg	Ser	Phe	Lys	Ile	
		610					615				620						
	Gly	Asp	Asn	Val	Val	Ser	Val	Asn	Ser	Ile	Val	Glu	Ile	Asn	Ser	Ile	
	625					630				635					640		
	Lys	Gln	Gly	Glu	Tyr	Lys	Ile	Thr	Pro	Ile	Asp	Asp	Lys	Ala	Gln	Phe	
				645						650					655		
	Tyr	Ile	Phe	Tyr	Leu	Lys	Asp	Ser	Ala	Ile	Pro	Tyr	Ala	Gln	Phe	Ile	
			660						665					670			
	Leu	Met	Asp	Lys	Thr	Met	Phe	Asn	Ser	Ala	Tyr	Val	Gln	Met	Phe	Phe	
			675				680						685				
	Leu	Gly	Asn	Tyr	Asp	Lys	Asn	Leu	Phe	Asp	Leu	Val	Ile	Asn	Ser	Arg	
		690					695					700					
	Asp	Ala	Lys	Val	Phe	Lys	Leu	Lys	Ile	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	
	705					710				715					720		

Tyr Ala

<210> 27
 <211> 8171
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列描述: 合成
 多核苷酸

[0028]

<400> 27
 gcggccgcaa ggggttcgcg tcagcgggtg ttggcgggtg tcggggctgg cttactatg 60
 cggcatcaga gcagattgta ctgagagtgc accatatgcg gtgtgaaata cgcacagat 120
 gcgtaaggag aaaataccgc atcaggcgcc attcgccatt cagctgcgca actgttggga 180
 agggcgatcg gtggggcctt ctctctattt acgccagctg gcgaaagggg gatgtgctgc 240
 aaggcgatta agttgggtaa cgccagggtt ttcccagtcg cgacgttgta aaacgacggc 300
 cagtgaattg taatacgact cactataggg cgaattcgag ctgggtaccc ggggatccca 360
 cgtggcgcg cactagtgtc agcgacgtcg tgggacccgc tagagtcgac ctgcaggcat 420
 gcaagcttga gtattctata gtctaccta aatagcttgg cgtatcatg gtcatactg 480
 ttctctgtgt gaaattgta tccgtctaca attccacaca acatacgcgc cggaagcata 540
 aagtgtaaag cctgggggtg ctaatgagt agctaactca cattaattgc gttgcgtcga 600
 ctgcccgtt tccagtcggg aaacctgtcg tgccagctgc attaatgaat cgcccaacgc 660
 gaaccccttg cgccgcgcg ggcgctcgac caattctcat gttgacgc ttatcatcga 720
 atttctgcca ttcatcgtt tattatcact tatteaggcg tagcaaccag gcgtttaagg 780
 gcaccaataa ctgccttaaa aaaattacgc cccgcctgc cactcctgc agtactgttg 840
 taattcatta agcattctgc cgacatggaa gccatcaca acggcatgat gaacctgaat 900
 cgccagcggc atcagacct gtctgccttg cgtataatat ttgcccattg tgaacacggg 960
 ggcaagaag ttgtccatat tggccaggtt taaatcaaaa ctggtgaaac tcacccaggg 1020
 attggtctag acaaaaaaca tattctcaat aaaccttita gggaaatagg ccaggttttc 1080
 accgtaacac gccacatctt gcgaatata gtgtagaaac tgccggaaat cgtcgttgta 1140
 ttactccag agcgatgaaa acgtttcagt ttgtcatggt aaaacgggtg aacaagggtg 1200
 aacactatcc catatcaca gctcaccgtc ttctattgcc atacgaaatt cggatgagc 1260
 attcatcagg cgggcaagaa tgtgaataaa ggccggataa aacttgctgt tatttttttt 1320
 tacgggtctt aaaaaggcgc taatatecag ctgaacggtc tggttatagg tacattgagc 1380
 aactgaactg aatgcctcaa aatgtttttt acgatgccat tgggatata caacgggtgt 1440
 atatccagtg atttttttct ccatttttag ttcttttagt ccgaaaaatc tcgataactc 1500
 aaaaaatacg cccggtagt atcttatttc attatggtga aagttggaac ctcttaactg 1560
 ccgatcaacg tctcatcttc gccaaaagtt ggcccagggc ttcccgttat caacagggac 1620
 accaggattt atttattctg cgaagtgate ttccgtcaca ggtatttatt cgcgataagc 1680
 tcatggagcg gcgtaacgtc cgcacaggaa ggacagagaa agcgcggatc tgggaagtga 1740
 cggacagaac ggtcaggacc tggattgggg aggcgggttc cgcgcgtgct gctgacggtg 1800
 tgaegtcttc tgttccggtc acaccacata cgttccgcca ttcttatgcg atgcacatgc 1860

[0029]

tgatgcccgg tataccgctg aaagtctctg aaagcctgat gggacataag tccatcagtt	1920
caacggaagt ctacacgaag gtttttgcgc tggatgtggo tgcceggcac cgggtgcagt	1980
ttgcgatgcc ggagtcigat gcggttgcga tgcctgaaaca attatectga gaataaatgc	2040
cttggccctt ataatggaaat gtggaaciga gttgatatgc tgtttttgtc tgttaaacag	2100
agaagctggc tgttatccac tgagaagcga acgaaacagt cgggaaaate tcccattatc	2160
gtagagatcc gcattattaa tctcaggagc ctgtgtageg ttatataggaa gtagtgttct	2220
gtcatgatgc ctgcaagcgg taacgaaaac gatttgaata tgccttcagg aacaatagaa	2280
atcttcgigc ggtgttaact tgaagtggag cggattatgt cagcaatgga cagaacaacc	2340
taataaacac agaaccatga tgtggctctg ctttttacag ccagttagtc tgcgcgcagt	2400
cgagcgacag ggcgaagccc tcgcttggtt gccctgcgcg ctgggctggc ggcctctat	2460
ggccctgcaa acgcgcaga aacgcctgcg aagccgtgtg cgagacaccg cggccggccg	2520
ccggcgttgt ggataacctg cggaaaactt ggcctcact gacagatgag gggcggacgt	2580
tgacacttga ggggcgcact caccgcgcgc ggcgttgaca gatgaggggc aggcctcatt	2640
tcggccggcg acgtggagct ggcgcgcctc gcaaatcggc gaaaacgcct gattttacgc	2700
gagtttccca cagatgatgt ggacaagcct ggggataagt gccctgcggt attgacactt	2760
gaggggcgcg actactgaca gatgaggggc gcgctccttg acacttgagg ggcagagtgc	2820
tgacagatga ggggcgcacc tattgacatt tgaggggctg tccacaggca gaaaatccag	2880
catttgcag ggtttccgcc cgtttttcgg ccaccgctaa cctgtctttt aacctgcttt	2940
taaaccaata ttataaacc ttgtttttaa ccaggcctgc gccctgtgcg cgtgaccgcg	3000
cacgcggaag ggggggtgcc ccccttcctg aaccctcccg gtccagtgag cgaggaagca	3060
ccagggaaca gaacttatat attctgetta cacaagatgc ctgaaaaaac ttcccttggg	3120
gttatecaat tatecacggg gatattttta taattatttt ttttatagtt tttagatctt	3180
cttttttaga ggcctttgta ggcctttatc catgctggtt ctagagaagg tgttgtgaca	3240
aattgccctt tcagtgtgac aaatcaccct caaatgacag tccgtctctg gacaaattgc	3300
cettaaccct gtgacaaati gccctcagaa gaagctgttt ttccacaaag ttatccctgc	3360
ttattgactc ttttttattt agtgtgacaa tctaaaaact tgtcacactt cacatggatc	3420
tgtcatggcg gaacacggc ttatcaatca caagaaactt aaaaatagcc cgcgaatcgt	3480
ccagtcaaac gacctcactg aggcggcata tagtctctcc cgggatcaaa aacgtatgct	3540
gtatctgttc gttgaccaga tcagaaaate tgatggcacc ctacaggaac atgacggtat	3600
ctgcgagate catgttgcta aatatgcga aatattegga ttgacctctg cggaaagccag	3660
taaggatata cggcaggcat tgaagagttt cgcggggaag gaagtggttt tttatcgcgc	3720
tgaagaggat gccggcgatg aaaaaggcta tgaatctttt ccttggttta tcaaacgtgc	3780
gcacagtcca tccagagggc ttacacagtgt acatatcaac ccatatctca ttectttctt	3840
tatcgggcta cagaaccggt ttacgcagtt tcggcttagt gaaacaaaag aaatcaccaa	3900
tcctgatgcc atgcgtttat acgaatccct gtgtcagtat cgtaagccgg atggetcagg	3960
catcgtctct ctgaaaatcg actggateat agagcgttac cagctgcctc aaagttacca	4020
gcgtatgcct gacttccgcc gccgcttccct gcaggtctgt gttaatgaga tcaacagcag	4080
aactccaatg cgcctctcat acattagaaa aaagaaaggc cgcagacga ctcatatcgt	4140

[0030]

```

attttccttc cgcgatatca cttccatgac gacaggatag tctgagggtt atctgtcaca 4200
gatttgaggg tggttcgtca catttgittc gacctactga gggtaatttg tcacagtttt 4260
gctgtttcct tcagccctgca tggattttct catacttttt gaactgtaat ttttaaggaa 4320
gccaaatttg agggcagttt gtcacagtig atttcccttc ctttcccttc gtcagtigac 4380
ctgatatcgg gggttagttc gtcacattg atgagggttg attatcacag ttattactc 4440
tgaattgctt atccgcgtgt gtacctctac ctggagtttt tcccacggtg gatatttctt 4500
cttgcgtcga gcgtaagage tatctgacag aacagttctt ctttgccttc tcgccagttc 4560
gctcgcatac ctgcgttaca cggeicgggc gagecctagt gataataagt gactgaggta 4620
tgtctcttc tttctctctt ttgtagtgtt gctcttattt taaacaactt tgcgggtttt 4680
tgatgaactt gcgattttgt tgttgccttg cagtaaatg caagatttaa taaaaaacg 4740
caaagcaatg attaaaggat gttcagaatg aaactcatgg aaacacttaa ccagtgcata 4800
aacgcctggc atgaaatgac gaagcctatc gccattgcac agtttaatga tgacagcccg 4860
gaagcgagga aaataaccgg gcctgggaga ataggtagaag cagcggattt agttgggggt 4920
ttctctcagg ctatcagaga tgcgagaaa gcaggcgac taccgcacc ggataaggaa 4980
attcgaggac gggttgagca acgtgttggt tatacaattg aacaaattaa tcatacgtt 5040
gatgtgtttg gtaecgattt gcgacgtgtt gaagacgtat ttccaccgtt gatcggggtt 5100
gttgcctata aagggtggct ttacaaaacc tcagtttctg ttcattctgc tcaggatctg 5160
gctctgaagg ggtacgtgtt ttgtctctg gaagtaacg acccccagg aacagcctca 5220
atgtatcacg gatgggiacc agatcttcat attcatgcag aagacactct cctgccttc 5280
tatcttgggg aaaaggacga tgtcacttat gcaataaagc ccacttgctg gcgggggtt 5340
gacattatc cttctgtctt ggtctgcac cgtattgaaa ctgagttaat gggcaattt 5400
gatgaaggta aactgcctac cgtaccacac ctgatgtctt gactggccat tgaacgtt 5460
gtcatgactt atgatgtcat agttattgac agegcgcta acctgggtat cggeacgatt 5520
aatgttgtat gtctctcga tgtctgtatt gtccacacg ctgtcagttt gtttgactac 5580
acctccgcac tgcagttttt cgatatgett cgtgatctgc tcaagaacgt tgatcttaaa 5640
gggttcgagc ctgatgtacg tattttgett accaaataca gcaatagtaa tgctctcag 5700
tcccctgga tggaggagca aattcgggat gcctggggaa gcaigtctt aaaaaatgtt 5760
gtacgtgaaa cggatgaagt tggtaaaggt cagatccgga tgagaactgt ttttgaacag 5820
gccattgac aacgtctctt aactgggtgc tggagaaatg ctctttctat ttgggaacct 5880
gtctgcaatg aaattttcga tctctgtatt aaaccacgt gggagattag ataataagc 5940
gtgcgcctgt tattccaaaa catagctca atactcaacc ggttgaagat acttcgttat 6000
cgacaccagc tgcctcgatg gtggattcgt taattgcgag cgtaggagta atggtctcgc 6060
gtaatgceat tactttgctt gtatgtggtc gggatgtgaa gtttactctt gaagtctcc 6120
gggtgatag tgttgagaag acctctcggg tatggtcagg taatgaacgt gaccaggagc 6180
tgcttactga ggacgcactg gatgatctca tcccttctt tctactgact ggtaacaga 6240
caccggcgtt cggtcgaaga gtatctggtg tcatagaaat tgccgatgg agtcgcctc 6300
gtaaagctgc tgcacttacc gaaagtgtt atcgtgttct ggttggcgag ctggatgatg 6360
agcagatggc tgcattatcc agattgggia acgattatcg cccaacaagt gcttatgaac 6420

```

	gtggtcageg ttaatgaagc cgattgcaga atgaatttgc tggaaatatt tctgcgctgg	6480
	ctgatgcgga aaatatttca cgtaagatta ttaccectg tateaacacc gccaaattgc	6540
	ctaaatcagt tgttgccttt ttttctcacc cgggtgaact atctgcccg ttaggtgatg	6600
	catttcaaaa agccttttaca gataaagagg aattacttaa gcagcaggca tctaaccttc	6660
	atgagcagaa aaaagctggg gtgatatttg aagctgaaga agttateact cttttaactt	6720
	ctgtgcttaa aacgtcatct gcatcaagaa ctagttaaag ctacagacat cagtttgctc	6780
	ctggagcgac agtattgtat aaggcgata aaatggctgt taacctggac aggtctcgtg	6840
	ttccaactga gtgtatagag aaaattgagg ccattcttaa ggaacttgaa aagccagcac	6900
	ccgatgcga ccacgtttta gtctacgttt atctgtcttt acttaatgtc ctttgttaca	6960
	ggccagaaag cataactggc ctgaatatc tctctgggcc cactgttcca ctgtatctgt	7020
	cggctctgata atcagactgg gaccacggtc ccaactgtat cgtcggctctg attattagtc	7080
	tgggaccacg gtcccaactg tategtcggc ctgattatta gtctgggacc acggtccca	7140
	tcgtatctgc ggtctgataa tcagactggg accacggctc cactcgtate gtcggctctga	7200
	ttattagctt gggaccatgg tcccactcgt atcgtcggtc tgattattag tctgggacca	7260
	cggctccact cgtatcgtgc gtctgattat tagtctggaa ccacggctcc actcgtatcg	7320
	tcggctctgat tattagctgc ggaccacggc cccactcgt tctcggctct gattattagt	7380
	ctgggaccac gateccactc gtgttgctgg tctgattate ggtctgggac caccgtccca	7440
	cttgtattgt cgatcagact atcagcgtga gactacgatt ccatcaatgc ctgtcaaggg	7500
[0031]	caagtattga catgtcgtgc taaccigtat aacggagtaa cctcgggtgtg cggttgtatg	7560
	ccgtcgttgg attgctgcig tgtctgcctt atccacaaca ttttgcgcac gggtatgttg	7620
	acaaaatacc tggttaccac ggccgtgcgc gcacgttaac cgggctgcat ccgatgcaag	7680
	tgtgtcgtgc tcgacgagct cgcgagctgc gacatgaggt tgcgccgtat tcagtgtcgc	7740
	tgatttgtat tgtctgaagt tgtttttacg itaagtigat gcagatcaat taatacgata	7800
	ccctgcgtcat aattgattat ttgacgtggc ttgatggcct ccaacgacgt tgtgataatg	7860
	agatgataat cattatcact ttaacgggtc tttccgggtg tccgacaggt tacggggcgg	7920
	cgacctgcgc ggttttctgc atttatgaaa attttccggc ttaagcgctt tccgttcttc	7980
	ttcgtcataa cttaatgttt ttatttaaaa taccctctga aaagaannga aacgacaggt	8040
	gctgaaagcg agctttttgg cctctgtcgt ttcctttctc tgtttttgtc cgtggaatga	8100
	acaatggaag tccgagctca tcgctaataa ctctgtatag catacattat acgaagtat	8160
	attcgatcca c	8171
	<210> 28	
	<211> 20982	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 人工序列描述: 合成多核苷酸	
	<400> 28	
	ctagcggcaa aacgtatgcc gggtgacctc tctgaatact ccgtcatcca gaccaaagaa	60
	ccgtcggatc gcaaggttaa agtcagccgc attgttgaat ttatcgaaaa accggatcag	120

[0032]

```

ccgcagacgc tggactcaga catcatggcc gttggctcgt atgtgctttc tgcgatatatt 180
tgcccggaac ttgaacgtac tcagcctggt gcatggggac gtattcagct gactgaigcc 240
attgccgagc tggcgaaaaa acagtccgtt gatgcaatgc tgatgaccgg cgacagctac 300
gactgcggta aaaaaatggg ctatatgcag gcgtttgtga agtatgggct gcgcaacctg 360
aaagaagggg cgaagtccg taaaggatatt gagaagcgtt taagcgaata atgaaaaatct 420
gaccggatgt aacggttgat aagaaaatta taacggcagt gaagattcgt ggtgaaagta 480
atttgttgcc aatatccctg ccgttgtttt atataaaca tcagaataac aacgagttag 540
caataggatt ttagtcaaag tttccagga ttttccttgt ttccagagcg gattggtaag 600
acaattagct ttgaatttt tcgggtttag cgcgagtggg taacgctcgt cacatcgtag 660
gcatgcatgc agtgcctcgg tagctgtaaa gccaggggcg gtagegtgca ttaatacttc 720
tattaatcaa acigagagcc gcttattca cagcatgctc tgaagcaata tggaaataat 780
taggtgaaaa tacttgttac tggtggecga ggatttattg gttttgctgt agttcgtcac 840
attataaata atacgcagga tagtgttgtt aatgtcgata aattaacgta cgcggaaac 900
ctggaatcac ttctgatgt ttctgattct gaacgctatg ttttgaaca tgcggatatt 960
tgcgatgcag ctgcaatggc acggattttt gctcagcatc agccagatgc agtgatgcac 1020
ctggctgctg aaagccatgt tgaccgttca attacaggtc ctgcggcatt tattgaaacc 1080
aatattgttg gtacatatgt ccttttgga gccgctcga attattggtc tgcctttgat 1140
agcgacaaga aaactagatt ccgttttcat catatttcta ctgacgaagt ctatggtgat 1200
ttgccctate ctgacgaggt aaataatata gaagaattac ccttattac agagacaaca 1260
gcttacgcgc caagcagccc ttattccgct tcaaaagcat ccagcgatca tttagtccgc 1320
gggttgaaac gtacctatgg ttaccaacc atgtgacta attgctctaa taattatggt 1380
ccttateatt tcccgaaaaa attgattcca ttggttatct tgaatgctct ggaaggtatg 1440
gcattacta tttatggcaa aggggatcaa attcgtgact ggcgtgatgt tgaagatcat 1500
gcgctgctgt tatataccgt cgtaaccgaa ggtaaagcgg gtgaaactta taacattggt 1560
ggacacaaac aaaagaaaaa catcgatgta gtgctacta tttgtgattt gctggatgag 1620
attgtaccga aagagaaatc ttaccgegag caaattactt atgttgccga tcgcccggga 1680
cacgatgcgc gttatgcgat tgatgcagag aagattagcc gcgaattggg ctggaaacgg 1740
caggaaacgt ttgagagcgg gattcggaag acattggaat ggtacctgtc caatacaaaa 1800
tgggttgata atgtgaaaag tgggtcttat caatcgtgga ttgaacagaa ctatgagggc 1860
cgccagtaat gaatatectc ctttcggca aaacagggca ggtagggttg gaactacagc 1920
gtgctctggc accittgggt aatttgattg ctcttgatgt tcaactcaat gattattgtg 1980
gtgattttag taatcctgaa ggtgtagctg aaacagtcaa aagaattcga cctgaigtta 2040
ttgttaatgc tgcggctcac accgcagtag ataaggctga gtcagaaccc gaatttgcac 2100
aattactcaa tgcgactagt gttgaatcaa ttgcaaaaga ggctaatgaa gttggggett 2160
gggtaattca ttactcaact gactacgtat tccctgaaa tggcgacacg ccattggctg 2220
agacggatgc aaccgcaccg ctaaatgttt acggtgaaac caagttagcc ggagaaaaag 2280
cgttacagga acattgcgcg aagcatctta ttttcgtac cagctgggta tacgcageta 2340
aaggaaaata cttcgccaaa acgatgttgc gtctggcaaa agagcgcgaa gaactggctg 2400

```

[0033]

tgataaatga tcaatttggg gegccaacag gtgtgagct gctggctgat tgtaaggcac	2460
atgctatteg tgtggcactg aataaacagg aagtcgcagg ttgtaccat ctggtagcca	2520
gfgtgaccac aacctggcac gattatgctg cgttggtttt tgaagaggcg cgaaaageag	2580
gtattcccc tgcactcaac aagctcaacg cagtaccaac aacagccat cctacaccag	2640
ctcgtcgtec acataactct cgccttaata cagaaaaaft tcagcagaac tttgcgcttg	2700
tcttgcciga ctggcagggt ggtgtgaaac gaatgctca cgaattaatt acgactacag	2760
caatttaata gtttttgcgt cttgttcgtg atggtggagg aagatgaatt aaaaggaaatg	2820
atgaaatgaa aacgcglaaa ggtattatit tagcgggtgg ttctggtaca cgtctttatc	2880
ctgtgactat ggctgtcagt aaacagctat tactattia tgataagccg atgatctatt	2940
acccgcctct tacactgatg ttggcgggta ttccgatat tctgattatt agtacgccac	3000
aggatactcc tcgttttcaa caactgciag gtgacggtag ccagtggggg ctaaattctc	3060
agtacaaagt gcaaccgact ccagatgggc ttgcgcaggc gtttattatc ggtgaagagt	3120
ttatcggtgg tgatgattgt gctttggctt ttggtgataa tatctctac ggtcatgac	3180
tgccgaagtt aatggatgct gctgttaaca aagaaagtgg tgcaacggta tttgectatc	3240
acgttaatga tccigaacgc taecgcgtcg ttgagtttga taaaaacggc acggcaataa	3300
gcctggaaga aaaaccgcta caaccaaaaa gtaattatgc ggtaaccggg ctttatttct	3360
atgataacga cgttgtcgaa atggcgaaaa accttaagcc ttctgccgt ggtgaactgg	3420
aaattaccga tattaaccgt atttatatgg aacagggcg tttatccgt gccatgatgg	3480
ggcgtgggta tgcattggcg gatacgggga cacatcagag tcttattgaa gcaagcaact	3540
tcattgccac cattgaagag cgccaggac taaaggtttc ctgccagaa gaaattgett	3600
accgtaaaagg gtttattgat gctgaacagg tgaaagcatt agcggagccg ctgaaaaaaa	3660
atgettiatgg acagtatctg ctgaaaaatg ttaaaggtta ttaataaaat gaacgtaatt	3720
aaaacagaaa ttctgatgt gtttaatttc gagccgaaag tttttggtga tgagcgtggt	3780
ttctttatgg aaagctttta tcagaaagtt ttcaagaag ctgtaggacg taaggttgaa	3840
ttgttcagg ataaccattc gaagcttagt aaagggtttt taecgggget gcattatcag	3900
ttagaacctt atgcgcaagg gaaactggta cgttgcttg ttggtgaggt tttgatgta	3960
gctgttgata ttctgaaac gtgcctacc ttigttaaaf gggttggggt gaatttatct	4020
gctgagaata agcggcaatt gggatccct gagggatttg cacatggitt tttggtgctg	4080
agcgagactg cggaattttt atataaaacg acgaactatt atcatctga tagtgataga	4140
gggattgtat ggaatgatcc tattcigagc ataaaatgac cgacgataga acataataat	4200
tatattttat cgattaaaga tgcaaggget aaagaattgc ataacatgaa ggaattattt	4260
ttgtgagtat tgtaagaat acittatgga atataagtgg gtataattata ccatcattaa	4320
tagcaattec tgcgttaggt atactgtct gaattctagg gaccgagcaa tttggccttt	4380
ttacgttagc tattgcctta gttggatatg caagtatttt tgatgctgga ttgaccagag	4440
ctgttataag agaagtatca atalataaaa atgttcataa agaattlaaga gcatcattt	4500
caacttcaac ggtaattcta actatattgg gcttgattgg cggtagtgta ctatttttga	4560
gtagcaatgt aattgttaaa ttattaaaca ttaacgcgaa tcatgttgta gaatctgtca	4620
aagcaatata tattatttca gctaccatac cttataactt gttaaacca gctcgggttg	4680

[0034]

```

ggatTTTTga ggggatggaa aagttcagaa aagtaaattt aataaaatca attaacaact 4740
cttttTgtgc tggattacca gtgattttct gtttttttca tggaggatta ctaagtgcta 4800
tatatggttt agttatggca agagtcttat cacttatagt gacctttata tttagtcgaa 4860
aactaataat atcatctggg ctgtctgtaa aaattgtaac agttaaaaga ttaatcggct 4920
ttggaagctg gataacagtt agcaatatta ttageccctat tatgacatat atggatcggt 4980
ttattctttc acacattgtg ggggcTgata aagtttcttt ttatactgct cegtcTgaag 5040
gtatacaacg cttaacgata ttaccaagtg cgttgTccag agctattttt ccaagattaa 5100
gttcagaatt gcaatcggta aagcaaaact aaatattatc atattttata atggttatTg 5160
gtatacttcc aattgtaatg ttgataatta ttttatcaga ttttataatg tccgctTgga 5220
tgggacctac atatcatggg acgccaggta tagtattaaa aattctTgca ataggtttct 5280
tttttaattg catTgcacaa atcccatTg tttcagtTca ggctagtTga agaTcaaaaa 5340
ttacagctat tattcatTg ctTgaagTta tcccatattt atgcataTta tatattttta 5400
tttatcatTg gggaattgtt ggagccgcaa tagcatggTc tGtaagaaca tegttagatt 5460
ttttgatatt attattaatt gatacgaaat attaatagcg aattgatttt agggattact 5520
tcctcaagcc catctaatTa gagtGcaaac atgacttctg atttttataa ctcaaaagac 5580
aaaagtTtaa gtgttctttt gttttttggg tttatatTtt tcctTcacg tagctttcca 5640
tttattcaat atagtTggat tatggagggg tttttatgtc ttgtatcat gtcattTaca 5700
aagaaaattg caaacggaat atatcactat cctgttatTt taatatTctt attagctctt 5760
tttataaatt ttatttatTc ctatatcaag ggtaacgata tagcgataat aattaggttt 5820
tatattatca tattatttat attatgtctt tattctctgt ctlatTgaac catctcgatt 5880
gttaaaatat ttttatattt aatggtatta cagcgggtta ttataTccat cattagtatt 5940
tatatgacaa aaacatatgg tattggtgat tattcagcac taagacatta tttttTggag 6000
aatgattatg gtgatgttta tacatatTga agtggtttct atagagtTca aattaaagga 6060
aatgctctca ttccattTgc ctttatgtTg catatagtca taaaagatta ttctattat 6120
cgattcaaaa atacaataac cgttatctTg gctataggta ctatagtTgc tggtaatttt 6180
gcataattTg tttegatatg cttgtttttt atgtatatta tactatgtTc taaatetaac 6240
tcacgatacg ctaaattaag gaaaattatt ttTggggTtt ttcttactgt gattctccct 6300
ttttttatta catattcaat tgagtTgata atcatgaaat caaatggagc tgattcttct 6360
ttaggagtta gatgggatca gtttactgta ttaattaatg atctTcacaga gtcTgtatca 6420
aattttgTta taggttctTg ttTgggTaat gTcatcaaaa ttcaaaactcc tatccTgat 6480
tatagtcat atatatatta tgaattgcag tcagtTtatt ttttaaatca actTgcggtt 6540
attttatTta ctttgttttt attaattaat ctctctctca cgattaaaat cataaaatac 6600
agtgagtTgt gtgtctata tttctatfat gtttcttatg caattactaa tcttatatt 6660
ttagactcta accatgtTgc tGtaataatt gtattagtga catTaaagtaa tgtTctaaaa 6720
aagatgaaag ctaaataag gtttaaggt gaagatggac acTgtatTg cgtttTggt 6780
tgctTacaac ccagaacata atgattTaaa aaatgcggtt gaattattgt tgagacaagt 6840
tactaaagtt gTcgtTtTga ataactctac aaatggttat aaatatgctg aaaattcttc 6900
agcgatgta aaaaattTca attTcaatga taatttaggc atagcagaag cccaaagtat 6960

```

[0035]

```

aggaatgaaa tgggccttttg aaaatggcgc tgattttata ttgcaaatgg atcaggatag 7020
tattccctgat cctaaagatgg tagagcagtt acttacttgt tacaaaaaat tgcctaaaca 7080
aaatgtcaat gttggittag ttggttcaca agattttgat aaagtaactg gtgaattaaa 7140
taaagcaagg gtaaaaaaag ggaaaccact tacagaagtt tattatgagg tagatagtac 7200
attaagtctt ggcagctcaa taccaaaaaa tagttggttg attgttggag gaatgaaaga 7260
tgagcttttt atcgatgcgg tagaccatga atattgttgg agattaagag ctgctgggtt 7320
taaagtaatt aggaataaaa atgcgttact tgcacataga ctgggagatg ggcgatttaa 7380
gatcttaaat attctttctg tcggtttgcg aagcccatth cgtcattatt atgetactcg 7440
aaatatcttt cttttattaa ataaaaatta tgtaaccatc tactggaaaa ttctagtct 7500
ggttaaatta attggaaagg tttttttata tectattttc ctcccaaatg gtaataaaag 7560
gttatatttt tttttaaaag gcattaatga cggtttaatg ggtcgaagtg gtaaaaigaa 7620
atgaatcata gattagaaaa attctcagtt ttaattagca ttataaaaa tgatctaccg 7680
caatittttg aggtggctct acgccttatt ttccacgac aaacacttaa gccagatcaa 7740
atagtaattg ttgcagatgg agaactccat caaacacaca tcgataattat aaattcattc 7800
attgatgatg ttggcaataa aatagtaaca ttgtaccctt taccatagaaa tgttggattg 7860
gctaatgcct taaatgaagg attaaagct ttaggaatg agttagtggc aagaatggat 7920
gcgatgata ttcttttgcg tcacgggttt gagaacaaa ttctttttat gattaataat 7980
tcagaaatag atgtatgtgg cagttttatt gatgaaattg aaactgttac tgaggagttt 8040
atttcaacac gcaaagtgcc tctcgaacat agagaaatag ttaaattcgc gaggaacga 8100
agcgcagtta gccatccttc tgtaattttt agaaagaata cagtattagc tgttgggtgt 8160
tatctctcat tcagaaaatc tcaagatttt gcattgtgga gcttattaat tgiacataat 8220
gcaagatttg caaatcttcc agatatttta ttaaaaatgc gaactggctg taatcttatg 8280
gctcgacgtg gattgtcata tttattgtac gagtataaag tattgtattt tcaatataaa 8340
attggtttta ttcgaaaaaa tgaattaata agtaatgcta tgttgagaac attttttctg 8400
ataatgccat ctaaattaaa ggagctgatg tattcaatcg ttaggaateg ataataataa 8460
ttttctgatt aagtgttatg gatttatttt tattaggeat attctataat taagcataac 8520
ccgcatacca ccagcggta tctgacagg agtaacaaat gtcaaagcaa cagatcggcg 8580
tcgtcggtat ggcagtgatg gggcgcaacc ttgcgtcaa tatcgaaagc cgtgggtata 8640
ccgtctctat tttaacogt tcccgtagaa agaccgaaga agtgattacc gaaaatccag 8700
gcaagaaaact ggttccctac tatacgggtg aagaatttgt tgaatctctg gaaacgctc 8760
gtcgcatcct gttaatgggt aaagcaggtg ctggcacgga tgetgetatt gattccctca 8820
agccatacct cgataaagggt gacatcatca ttgatgttgg taacaccttc ttcatgaca 8880
ccattcgctg taaccgtgag cttctcag aggcctttaa cttatcggt accggtgttt 8940
ccggtgggtg agaaggtgag ctgaaaggtc ctccattat gcctgggtggg cagaaagaag 9000
cttatgaact gattgcgccc atctgacca aaatcgccc tgtggctgaa gacggcgaa 9060
cgtcgcttac ctatattggt gccgatgtg caggtcatta tgtgaagatg gttcacaacg 9120
gtattgaata cggatgatg cagctgattg ctgaagccta ttctctgctt aaaggtggct 9180
tgaacctcac caacgaaga ctggcgaga cctttaccga gtggaataac ggtgaactga 9240

```

[0036]

```

gcagctacct gatcgacatc accaaagata tcttcaccaa aaaagatgaa gagggtaact 9300
acctggttga tigtattctg gatgaagcag caaacaaggg tacgggcaaa tggaccagcc 9360
agagcgcgct ggatcteggc gaaccgcigt cgttgattac cgagtctgtg ttgcacgtt 9420
atatctcttc tctgaaagag cagcgtgttg ccgcactctaa agttctctct ggcgccgaag 9480
cgcagccagc tggcgacaat gctgagtcca tegaaaaagt tcgcctgtcg ctgtatctgg 9540
gcaaaatcgt ttcttacgti cagggettct ctcagctacg cgttgcgtct gaagagtaca 9600
actgggatct gaactacggt gaaatcgaga agatttccg tctgtgtctg atcaccgtg 9660
cgcagttcct gcagaaaac accgatgctt atgccgaaaa tccgcagatc gctaacctgt 9720
tctgtgctcc ttacttcaag caaattgcg atgactacca gcaggcgtg cgcgatgtg 9780
tcgtttacgc agtacagaa ggtatcccg tgcgcacct cgcgcgtcg gtgtctatt 9840
acgacagcta ccgcgcgct gtctgcctg cgaacctgat ccaggcacag cgtgactatt 9900
tcggtgcgca tacttataag cgcattgata aagaagggt gtccatacc gaattgctgg 9960
attaatctga tttaaatcaa ttaatcaaa caagccccg agaaacctc cgggctttt 10020
tattatacaa ageggcaggt tagggcctt ttttataatt tatagttaa aacgcgatat 10080
aatacagcgc cgcacagcag gatcgtctcc ttgacagttc atctacatca ggttaaaaa 10140
tcccgcagta gatgaagctg tgggtgtgga ttaatgacca ctctaaatgt ttaaccggaa 10200
gaagtcagag ctaatgaaaa taacaattc aggaacaggt tatgttggtc ttcaaatgg 10260
tattctgatt gcgcaaaacc acgaagtgtg tgcactggat atcgttcagg ccaaatgga 10320
catgctaac aagaggcagt caccgcttgt tgataaggag attgaagagt atciggcgac 10380
taaagatctc aatttccgc clacgacaga taagtaagc cgtataaaa atgccgatta 10440
cgttattatt gccacacct ccgattaiga tccgaaaaca aattatttta atacctcaag 10500
cgtggaagcg gtcattcgtg atgtgacaga aattaatecc aacgcggtaa tgattataaa 10560
atcaactatc cctgttggtt ttacagagtc cattaaagaa cgttttggtt ttgaaaatgt 10620
gatcttttcg cctgagtttt tgcgtgaagg taaagcactt tatgataact tacaccatc 10680
acgcattgtg attggcgagc agtctgaacg cgtataacgt ttgtctcgt tattacagga 10740
aggcgccatt aagcaagaca taccaacatt gtttactgac tcaaccgagg ctgaggegat 10800
taaacttttt gcaaacactt atctggcgat gcgtgtagcg tatttcaatg aacttgatag 10860
ttatgctgaa agcctgggac ttaattcagc ccagattatt gagggcgat gccitgaccc 10920
gcgtatcggt aatcactaca acaaccgctc attcggttat ggtggttatt gtcgcccga 10980
agatactaag cagttaactg caaattacca gtctgtccg aataacctga tctcggeaat 11040
tgttgacgcc aaccgcagc gcaaagattt tattgccgat tctatccttg caegtaaac 11100
gaaagtgtgt ggcgtctatc gtgtgattat gaagaatggt tcagacaatt ttcgtgctc 11160
ctcgattcag ggtattatga agcgaatcaa ggccgaaagg gtgcctgtaa tegtattatga 11220
gccagctatg aaagaggacg atttttccg gtcgcgcgtg gtaectgate tggatgcgtt 11280
caaacaagaa gctgatgtta ttatttetaa ccgtatgtct gcgatctgg ctgatgtacg 11340
agataaagtt tataccgcgc acttgttttg caatgattaa ttatttgtt tcattctaag 11400
aaaaggccct aataaattag ggcccttctc tatggttttg taaaatcaaa ctttatagaa 11460
gttacgatac cattctacaa agttctttac cccttcttta acgacgttt caggtttgaa 11520

```

[0037]

tccatattacg tcatacagtg cttttgtatc agcactgggtt tccagtacat caccgggttg	11580
gagaggcacc atatttttgt tggcttcaat acccagagcc tcttctaacg cattgatata	11640
gtccatcaac tccacaggcg aactattacc aatggtatag acacgatatg gtgcigaact	11700
tgttgcaggc gagectgttt ctacagecca ctgtgggttt ttttctggaa taacatcctg	11760
taagcgaata atagcttcgg caataatcgc aatgtaagta aagtcacgct tcattttgcc	11820
gaagttgtaa acatcaatgc ttttaacctc cagcatggct ttagtgaatt taaataatgc	11880
catatccgga cgtecccatg gaccataaac cgtaaagaaa cgcagccctg tggtcggtaa	11940
gccatacaaa tgagaaiatg tatgggccat gagttcattc gcttttttag ttgctgcata	12000
aagcgaacaa ggaatgacta cagagtcacg tgiagagaaa ggcactctgc gggtcatgcc	12060
ataaacagaa ctggaggaag cgtaaagtag atgctgaaca ttattatggc gacatccttc	12120
tagtatgttc aggaatccaa tcaggtttgc atctgcatat gcattgggat tttcaagaga	12180
gtaacgtaca ccggcttgcg cagcgagggt tattacgggt tcgaaccgct cgtctgcaaa	12240
cagtgccgcc attttctcac gatcggccag gtcaatttta taaaaactga agttgtcgtg	12300
cttgagtaaa tcaagtcgtg ctgttttgag gttgacatcg taataatcat ttaagttgc	12360
aatgectaca accgatgac cagctgcaag aagccgttta cttagataga aaccgataaa	12420
gccagcagct ccgtaacca gaaatttcct ttataatcct cgtcaggct agaatatagc	12480
caatcttcct ctggcataac tgaagttaa attataccgt tagacaagaa aaaaagataa	12540
tcggtatcag ttctaaactt ggctgttttt tctggtaacg tgcctatttt acaatcaaag	12600
ctgttctaag ctgactatac aagccgacgt cattatctcc aaccgtatgg cagaagagct	12660
taaggatgtg gcagacaaa tctacaccgg cgatctcttt ggcagtgact aacatccgt	12720
tatcatggcg attttcgccc tgattctctt atgttccctt tgaataaatt cattattttt	12780
atcattttat ctatagcatt catggcgatt atcgctaaac tatggcgcg cgccaagtgg	12840
gatccccggg taccgagctc gaattcgccc tatagttagt cgtattacaa ttcactggcc	12900
gtcgttttac aacgtcgtga ctgggaaaac cctggcgtaa cccaacttaa tgccttgca	12960
gcacatcccc ctctcgccag ctggcgtaat agcgaagagg cccgcaccga tcgcccctcc	13020
caacagttgc gcagctgaat ggcaatggc gcctgatcg gtattttctc cttacgcac	13080
tgtgcggtat ttacacccg atatggtgca ctctcagtae aatctgctct gatgccgat	13140
agttaagcca gccccgacac ccgccaacac ccgtgacgc gaaccccttg cggccgcac	13200
gaatataact tcgtataatg tatgctatac gaagttattt gcgatgagct cggacttcca	13260
ttgttcattc caccgacaaa aacagagaaa ggaaacgaca gaggccaaaa agctcgtctt	13320
cagcacctgt cgtttctttt cttttcagag ggtattttta ataaaaacat taagttatga	13380
cgaagaagaa cggaaacgcc ttaaaccgga aaattttcat aaatagegaa aaccgcgag	13440
gtgcgccccc cgtaacctgt cggatcacgc gaaaggaccc gtaaagtgat aatgattatc	13500
atctacatai cacaacgtgc gtggaggcca tcaaaccacg tcaataaate aattatgaag	13560
caggatctgt attaatgat ctgcatcaac ttaacgtaaa aacaacttca gacaatacaa	13620
atcagcgaca ctgaatacgg ggcaacctca tgtccgagct cgcgagctcg tcgacagcga	13680
cacacttgca tcggatgcag cccggttaac gtgccggcac ggcctgggta accaggtatt	13740
ttgtccacat aaccgtgcgc aaaatgttgt ggataagcag gacacagcag caatccacag	13800

	caggcataca accgcacacc gaggttactc cgttctacag gttacgacga catgtcaata	13860
	cttgccttgg acaggeattg atggaatcgt agtctcaecg tgatagtctg atcgacaata	13920
	caagtgggac cgtgggtccc gaccgataat cagaccgaca acaegagtgg gatcgtggtc	13980
	ccagactaat aatcagaccg acgatacgag tgggaccgtg gtcccagact aataatcaga	14040
	ccgacgatac gagtgggacc gtggttccag actaataatc agaccgacga tacgagtggg	14100
	accgtggicc cagactaata atcagaccga cgatacgagt gggaccatgg tcccagacta	14160
	ataatcagac cgacgatacg agtgggaccg ttgtccagat ctgattatca gaccgacgat	14220
	acgagtggga ccgtgggtccc agactaataa tcagaccgac gatacgagtg ggaccgtggt	14280
	cccagactaa taatcagacc gacgatacga gtgggaccgt ggtcccagtc tgattatcag	14340
	accgacgata caagtggaac agtgggcccga gagagaatat tcaggccagt tatgctttct	14400
	ggcctgtaac aaaggacatt aagtaaagac agataaacgt agactaaaac gtggtcgcat	14460
	cagggtgctg gcttttcaag ttctttaaga atggcctcaa ttttctctat acactcagtt	14520
	ggaacacgag accgtgtccg gtttaagcacc attttategc cettatacaa tacgtcget	14580
	ccaggagcaa actgatgtcg tgagctttaa ctagttcttg atcgagatga cgttttaagc	14640
	acagaagtta aaagagtgat aacttcttca gcttcaaata tcaccccagc tttttctgc	14700
	tcatgaaggt tagatgcctg ctgcttaagt aattcctctt tatctgtaaa ggtttttga	14760
	agtgcacac ctgaccgggc agatagtcca ccgggggtgag aaaaagagc aacaactgat	14820
	ttaggcaatt tggcgggtgt gatacagcgg gtaalaatct tacgtgaaat atttccgca	14880
	tcagccagcg cagaaatatt tccagcaaat tcattctgca atcggttgc ataacgtga	14940
[0038]	ccacgttcat aagcacttgt tggcgataa tegttaacca atctggataa tgcagccatc	15000
	tgtctcatcat ccagctgccc aaaccagaaca cgataateac ttctggtaag tgcagcagct	15060
	ttacgacggc gactcccacg ggcattttct atgacaccag atactcttcg accgaacgcc	15120
	ggtgtctggt gaccagtcag tagaaaagaa gggatgagat catccagtgc gtcctcagta	15180
	agcagctcct ggtcaegtgc attacctgac catacccgag aggtcttctc aacactatca	15240
	ccccgggaca cttaagagt aaacttcaca tcccagaccac atacaggcaa agtaattggca	15300
	ttaccgcgag ccattactcc tacgcgcgca attaacgaat ccaccatcgg ggcagctggt	15360
	gtcgataaac aagtatcttc aaccggttga gtattgagcg tatgttttgg aataacagcg	15420
	gcacgcttca ttatctaate tcccagcgtg gtttaatcag acgatcgaaa atttcattgc	15480
	agacaggttc ccaaatagaa agagcatttc tccaggcacc agttgaagag cgttgataca	15540
	tggcctgttc aaaaacagtt ctcatccgga ctgaccttt accaacttca tccgtttcac	15600
	gtacaacatt ttttagaace atgttccccc aggcattccc aatttgetcc tccatccacg	15660
	gggaactgaga gccattacta ttgtgtatt tggtaagcaa aatacgtaca tcagctoga	15720
	accctttaag atcaacgttc ttgagcagat cacgaagcat atcgaaaaac tgcagtgcgg	15780
	aggtgtagtc aaacaactca gcaggcgtgg gaacaatcag cacatcagca gcacatacga	15840
	cattaatcgt gccgataccc aggttaggcg cgtgtcaat aactatgaca tcatagtcac	15900
	gagcaacagt ttcaatggcc agtcggagca tcagggtgtg atcggtgggc agtttacctt	15960
	catcaaattt gccattaac tcagtttcaa tacggtgcag agccagacag gaaggaataa	16020
	tgtaagccc cggccagcaa gtgggcttta ttgcataagt gacatcgtec ttttcccaa	16080

[0039]

gatagaaagg caggagagtg tcttctgcat gaatatgaag atcttggtacc catccgtgat 16140
acattgaggc tgttccctgg gggtcgttac cticcacgag caaaacacgt agccccitca 16200
gagccagatc ctgagcaaga tgaacagaaa ctgaggtttt gtaaacgcca cttttatggg 16260
cagcaacccc gatcacccgt ggaaatacgt cttcagcacg tcgcaatcgc gtaccaaaaca 16320
catcacgcat atgatttaatt tgttcaattg tataaccaac acgttgcctc acccgtctct 16380
gaatttccat atccgggtgc ggtagtcgcc ctgctttctc ggcatctctg atagcctgag 16440
aagaaaaccc aactaaatcc gctgtctcac ctattctcca gcgcggggtt attttectcg 16500
cttccgggct gtcacatta aactgtgcaa tggcgatagc cttcgtcatt tcatgaccag 16560
cgtttatgca ctggttaagt gtttccatga gtttcattct gaacatcctt taatcattgc 16620
tttgcgtttt ttatttaaat cttgcaattt actgcaaagc aacaacaaaa tcgcaaagtc 16680
atcaaaaaac cgcaaatgtg tttaaaaata gagcaacact acaaaaggag ataagaagag 16740
cacatacctc agtcacttat tatcactagc gctcgcgcga gcggtgtaac cgagcatagc 16800
gagcgaactg gcgaggaagc aaagaagaac tgttctgtca gatagctctt acgctcagcg 16860
caagaagaaa tatccaccgt gggaaaaact ccaggtagag gtacacacgc ggatagccaa 16920
ttcagagtaa taaactgtga taatcaaccc tcatcaatga tgacgaacta acccccgata 16980
tcaggtcaca tgacgaagg aaagagaagg aatcaactg tgacaaactg cectcaaatt 17040
tggttctctt aaaaattaca gttaaaaaag tatgagaaaa tccatgcagg ctgaaggaaa 17100
cagcaaaact gtgacaaatt accctcagta ggtcagaaca aatgtgacga accaccctca 17160
aatctgtgac agataaccct cagactatcc tgtcgtcatg gaagtgatat cgcggaagga 17220
aaatagata tgagtcgtct ggccgctttt ctttttctca atgtatgaga ggcgcattgg 17280
agttctgctg ttgattcctt taacacagac ctgcaggaag cggcggcgga agtcaggcat 17340
acgctggtaa ctttgaggca gctggtaacg ctctatgate cagtcgattt tcagagagac 17400
gatgcctgag ccacccggtt tacgatactg acacagggat tcgtataaac gcatggcata 17460
cggattggtg atttcttttg tttaactaag ccgaaactgc gtaaaccggt tctgtaaccc 17520
gataaagaag ggaatgagat atgggttgat atgtacactg taaagccctc tggatggact 17580
gtgcgcactg ttgataaacc aaggaaaaga ttcatagcct ttttcatcgc eggcactctc 17640
ttcaggcgca taaaaaacca cttctctccc cgcgaaactc ttcaatgcct gccgtatctc 17700
cttactggct tccgcagagg tcaatccgaa tatttcagca tatitagcaa catggatctc 17760
gcagataacc tcatgttctt gtaggggtgc atcagatttt ctgatctggt caacgaacag 17820
atacagcata cgtttttgat cccgggagag actatatgcc gccctcagtga ggctggttga 17880
ctggacgatt cgcgggctat ttttactgtt cttgtgattg ataaccgctg tttccgccat 17940
gacagatcca tgtgaagtgt gacaagtitt tagattgtca cactaaataa aaaagagtca 18000
ataagcaggg ataactttgt gaaaaaacag cttctctctga gggcaatttg tcacagggtt 18060
aagggaatt tgtcacagac aggactgtca tttaggggtg atttgtcaca ctgaaagggc 18120
aatttgteac aacaccttct ctagaaccag catggataaa ggctacaag gcctctataa 18180
aaagaagatc taaaaactat aaaaaaata attataaaaa tatccccgtg gataagtgga 18240
taacccaag ggaagttttt tcaggcatcg igtgtaagca gaatatataa gtctgttcc 18300
ctggtgtctc ctgcctact cgaccgggag ggttcgagaa gggggggcac ccccttcgg 18360

[0040]

```

cgtgcgcggt cagcgcaca gggcgagcc ctggttaaaa acaaggttta taaatatgg 18420
tttaaaagca ggttaaaaga caggtagcg gtggcgagaa aacggcgga aaccctigca 18480
aatgtcggat ttctgcctg tggacagccc ctcaaatgtc aatagggtcg cccctcatct 18540
gtcagcactc tgcctctcaa gtgtcaagga tcgcgcctct catctgtcag tagtcgcgc 18600
ctcaagtgt caataccgca gggcacattat ccccaggett gtccacatca tctgtgggaa 18660
actcgcgtaa aatcagcggt ttctgcgat ttgcaggett ggcagctcc acgtcgcgg 18720
ccgaaatcga gcctgcctct catctgtcaa cgcgcgcgg ggtgagtcgg cccctcaagt 18780
gtcaacgtcc gccctctate tgtcagttag ggccaagttt tcgcgaggt atccacaacg 18840
ccggcgcgcg gccgcgggt ctgcacacg gcttcgacgg cgtttctggc gcgtttgcag 18900
ggccatagac ggccgcagc ccagcgcgca ggccaaccag ccgagggett cgcctgtcg 18960
ctcgactcgg gcgagcacta ctggctgtaa aaggacagac cacatcatgg ttctgtgttc 19020
attaggttgt tctgtcatt gtgacataa tcgcctccac ttcaacgtaa caccgcacga 19080
agatttctat tgttctgaa ggcataatca aatcgttttc gttaccgett gcaggeatca 19140
tgacagaaca ctacttcta taaacgtac acagctcct gagattaata atgcggatct 19200
ctacgataat gggagatttt cccgactgtt tcgttcgett ctcagtggat aacagccagc 19260
ttctctgttt aacagacaaa aacagcatat ccactcagtt ccacatttc atataaaggc 19320
caaggcattt attctcagga taattgttc agcatcgcaa ccgcatcaga ctccggcatc 19380
gcaaacgtca cccgggtgcg ggcagccaca tccagcgcaa aaaccttcgt gtagacttc 19440
gttgaactga tggacttatg tccatcagg ctttgcagaa ctttcagcg tataccggca 19500
tacagcatgt gcacgcata ggaatggcg aacgtatgtg gtgtgaccgg aacagagaac 19560
gtcaaccggt cagcagcagc ggcggcaacc gcctcccaa tccaggtcct gaccgttctg 19620
tcgctcactt cccagatccg cgtttctct gtccctctctg tgcgacggtt acgcgcctcc 19680
atgagcttat cgcgaataaa tacctgtgac ggaagatcac ttgcagaat aataaatcc 19740
tggtgtccct gttgataccg ggaagccctg ggccaacttt tggcgaaaat gagacgttga 19800
tcggcacgta agaggttcca acttccacca taatgaaata agatactac cggcgctatt 19860
ttttgagtta tcgagatttt caggagctaa ggaagctaaa atggagaaaa aatcacctgg 19920
atataccacc gttgatatat cccaatggca tcgtaaagaa cattttgagg catttcagtc 19980
agtgtctcaa tgiacctata accagaccgt tcagctggat attacgcct ttttaaagac 20040
cgtaaagaaa aataagcaca agttttatcc ggcctttatt cacattcttg cccgcctgat 20100
gaatgcctat ccggaatttc gtaiggaat gaaagacggt gagctggtga taigggatag 20160
tgttcacct tgttacaccg ttttccatga gcaaacgtaa acgtttctat cgtctggag 20220
tgaataccac gacgatttcc ggcagtttct acacatatat tcgcaagatg tggcgigtta 20280
cggtgaaaac ctggcctatt tccctaaagg gtttattgag aatatgttt tcgtctcagc 20340
caatccctgg gtgagttcca ccagttttga tttaaacgtg gccaatatgg acaacttctt 20400
cgccccggtt ttcacctgg gcaaatatta tacgcaaggc gacaagggtc tgatgccgt 20460
ggcgattcag gttcatcatg ccgtttgtga tggcttccat gtcggcagaa tgcttaatga 20520
attacaacag tactgcgatg agtggcaggg cggggcgtaa ttttttaag gcagttattg 20580
gtgcccttaa acgcctggtt gctacgcctg aataagtat aataagcgga tgaatggcag 20640

```

	aaattcgatg ataagctgtc aaacatgaga attggctgac ggcccgggcg gccgcaaggg	20700
	gttcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccgac tggaaagcgg	20760
	gcagtgagcg caacgcaatt aatgtgagtt agctcactca ttaggcaccc cagcttttac	20820
	actttatgct tccggctcgt atgttgtgtg gaattgtgag cggataacaa ttccacacag	20880
	gaaacagcta tgacatgat tacgccaagc tatttaggtg agactataga ataactaagc	20940
	ttgcatgcct gcaggctgac tetagaggat cccacgacgt cg	20982
	<210> 29	
	<211> 22887	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 人工序列描述: 合成多核苷酸	
	<400> 29	
	ggggccgcaa ggggttcgcg tcagcgggtg ttggcgggtg tcggggctgg cttaactaig	60
	cggcatcaga gcagattgta ctgagagtc accatatgcg gtgtgaaata ccgcacagat	120
	gcgtaaggag aaaataccgc atcaggcgcc attcgccatt cagctgcga actgttggga	180
	agggcgatcg gtgcgggcct cttegetatt acgccagctg gcgaaagggg gatgtgctgc	240
	aaggegatta agttgggiaa cgccagggtt ticcagtcg cgacgttgta aaacgacggc	300
	cagtgaattg taatacgact cactataggg cgaattcgag ctcggtaccc ggggatccca	360
	cttggcgcgc cgccatagtt tagcgataat cgccatgaat gctataggat aaatgataaa	420
[0041]	ataatgaat tattacaaag ggaacataag agaatcaggc cgaaaatcgc catgataaca	480
	ggatgtagt cactgccaaa gagatcgcg gtgtagactt tgctgccac atccttaagc	540
	tcttctgcca tacggttga gataatgacg tcggcttgta tagtcagctt agaacagctt	600
	tgattgtaaa atgagcaegt taccagaaaa aacagccaag tttagaactg ataccgatta	660
	tcttttttct ttgtctaacg gtataattta actttcagtt atgccagaig aagattggct	720
	afattctage ctgagcgagg attataaatg aaatttcagg ttaacgggagc tgettgcttt	780
	atcgggttct atctaagtaa acggcttctt gcagctggtc atcaggttgt aggcattgac	840
	aaettaaatg attattacga tgtaacctc aaacaagcac gacttgattt actcaagcac	900
	gacaacttca gttttataaa aattgacctg gccgatcggt agaaaatggc ggcactgttt	960
	gcagacgagc gggtcgaacg cgtaataaac ctgcctgcgc aagccggtgt acgttactct	1020
	cttgaanaat ccaatgcata tgcagatgca aaectgattg gattcctgaa catactagaa	1080
	ggatgtgcgc ataataatgt tcagcatcta ctttacgctt cctccagttc tgtttatggc	1140
	atgaaccgca agatgccttt ctctacagat gactctgtag atcctcctgt ttcgctttat	1200
	gcagcaacta aaaaagcgaa tgaactcatg gcccatacat attctcattt gtatggctta	1260
	cggaccacag ggctgcgttt ctttacggtt tatggtecat ggggacgtcc ggatatgcca	1320
	ttattttaat tcaactaaagc catgctggaa ggtaaaagca ttgatgttta caacttcggc	1380
	aaaatgaagc gtgactttac ttacattgat gatattgccg aagetattat tcgcttacag	1440
	gatgttattc cagaaaaaaa cccacagtgg gctgtagaaa caggctcgcc tgcaacaagt	1500
	tcagcaccat atcgtgtcta taacattggt aatagttcgc ctgtggagtt gatggactat	1560
	atcaatgcgt tagaagagge tctgggtatt gaagccaaca aaaatatgat gcctctccaa	1620

	cccggtgatg tactggaac cagtgetgat acaaaagcac tgtatgacgt aataggattc	1680
	aaacctgaaa cgtcagttaa agaaggggta aagaactttg tagaatggta tcgtacttc	1740
	tataaagttt gatittacaa aaccataaga aaaggcccta attattagg gccttttctt	1800
	agaatgaaac aaaaataatta atcattgcca aacaagtcgc gcgtataaac ttatctgct	1860
	acatcagcca gatcggcaga catacggta gaaataataa catcagcttc ttgtttgaac	1920
	gcatecagat cactaccac gcgcgaccgg aaaaaatcgt cctctttcat agctggetca	1980
	taaacgattia caggcacacc ttctgccttg attcgcttca taataacctg aatcgaggaa	2040
	gcacgaaaaa tgcigaacc attcttcata atcaaacgat agacgccaac aactttcggt	2100
	ttacgtgcaa ggatagaate ggcaataaaa tctttgcgcg tgcggttggc gtcaacaatt	2160
	gccgagatca ggttatctgg cacagactgg taatttgcca gtaactgett agtatcttcc	2220
	ggcagacaat aaccaccata accgaatgac gggttgttgt agtgattacc gatacgcggg	2280
	tcaaggcata cgcctcaat aatctggcgt gaattaagtc ccaggcttcc agcataacta	2340
	tcaagttcat tgaataacgc tacacgcac gccagataag tgttcgcaaa aagtttaate	2400
	gcctcagcct cggttgagtc agtaaacat gttggtatgt ctgtcttaat ggcccttcc	2460
	tgtataacg cagcaaaacg tttagcgcgt tcagactgct cgccaateac aatgcgtgat	2520
	gggtgtaagt tatcataaag tgccttacct tcacgcaaaa actcaggcga aaagatcada	2580
	tttcaatac caaaacgttc ttaaatggac tctgtaaac caacagggat agttgatitt	2640
	ataatcatta ccgcgttggg attaatctct gtcacateac gaatgaccgc ttccacgett	2700
[0042]	gaggtattaa aataatttgt ttccgatac taatcggtag gtgtggcaat aataacgtaa	2760
	tcggcatitt tatacgcgtc atacttatct gtcgtagcgc ggaaattgag atctttagtc	2820
	gccagatact cttaactct cttatcaaca agcggtgact gcctcttgtt aagcatgtcc	2880
	actttggcct gaacgatate cagtgcaccc acttcgttgt tttagcgaat cagaatacca	2940
	tttgaaagac caacataacc tgttccfgaa attgttatti tcattagctc tgacttcttc	3000
	cggttaaaca tttagagtgg tcattaatcc accaccacag ctccatctac tgcgggattt	3060
	ttaacgtga tgtatagtaa ctgtcaaggc agcgatcctg ctgtgcggcg ctgtattata	3120
	tcgcgttttt aactataaat tataaaaaaa ggccctaacc tgccgctttg tataataaaa	3180
	aagcccgag ggittctcc ggccctgctt tgattaatg atttaaatca gattaatcca	3240
	gcatttcggt atggaacaca cttctttat caatgcgett ataagtatgc gcaaccgaaat	3300
	agtcacgctg tgcctggatc aggttcgcag gcagaacage ggccgcggtag ctgtcgtaat	3360
	aggcaaccgc agcggcgaag gtcggcaccg ggalaccgtt ctgtactgcg taagcgacga	3420
	catcgccag cgcctgcctg tagtcategg caatttgctt gaagtaagga gccagcaaca	3480
	ggttagcgat ctgcggattt tcggcataag catcggtgat ttcttcgagg aactgcgcac	3540
	ggatgatgca gccagcacgg aaaatcttcg cgatttcacc gtagttcaga tcccagttgt	3600
	actcttcaga cgcagcgcgt agctgagaga agccctgagc gtaagaaacg attttgccc	3660
	gatacagcgc acggcgaaat ttttcgatga actcagcatt gtcgccagct ggctgcgett	3720
	gggggcaga gagaatttta gatgcggcaa cagctgctc ttccagagaa gagatataac	3780
	gtgcaaacac agactcggta atcagcgaca gcggttcgcc gagatccagc gcgccttgge	3840
	tggctccatt gcccgtaact ttgtttgctg ctccatccag aatcacatca accaggtagt	3900

	taccctcttc atcttttttg gtgaagatat ctttggatgat gtcgatcagg tagctgctca	3960
	gttcaacggtt attccactcg gtaaaggctc gcgccagttc ttcgttgggtg aggttcaagc	4020
	cacctttaag cagagaaatag gcttcageaa tcagctgcat atcaccgtat tcaataccgt	4080
	tgtaaacat cttcacataa tgacctgcac catcggcacc aatataggta acgcacggtt	4140
	cgcctcttc agecacagcg gcgattttgg tcaggatcgg cgcaatcagt tcataagctt	4200
	ctttctgccc accaggcata atggaaggac ctttcagcgc accttcttca ccaccggaaa	4260
	caccgttacc gataaagita aagccttctg cagaaagctc acggttacga cgaatgggtg	4320
	catggaagaa ggtgttacca ccatcaatga tgatgtcacc ttatctgagg tatggcttga	4380
	gggaatcaat agcagcatcc gtgcacgac ctgctttcac cattaacagg atgcgacgag	4440
	gcgtttccag agattcaaca aattctttca ccgtatagta aggaaccagt ttcttgccctg	4500
	gattttcggg aatcacttct tcggctcttt caccggaacg gttagaaaata gagacggtat	4560
	aaccacggct ttcgatattg agcgcaagggt tgcgccecat cactgceata ccgacgacgc	4620
	cgatctgttg ctttgacatt gtttactect gtcaggatac cgctgggttg tatgcgggtt	4680
	atgcttaatt atagaatatg cctaataaaa ataatccat aacacttaat cagaaaatta	4740
	tiattatcga ttcttaacga ttgaatacat cagctccttt aatttagatg gcattatacg	4800
	aaaaaatgtt ctcaacatag cattacttat taattcattt ttctgaataa aaccaatttt	4860
	atattgataa tacaatactt tatactcgta caataaatat gacaatecac gtcgagccat	4920
	aagattacga ccagttcgca tttttaataa aatatctgga agatttgcaa atcttgcat	4980
[0043]	atgtacaatt aataggetcc acaatgcaaa atcttgagat ttcttgaatg gaggataacc	5040
	accaacagct aatactgtat tctttctaaa aattacagaa ggatggctaa ctgcgcttcg	5100
	ttctctcgcg aatttaacta ttctctatg ttcgagagge accttgcggtg ttgaaataaa	5160
	ctcttcagta acagtttcaa ttteatcaat aaaactgcca catacateta ttcttgaatt	5220
	attaatcata aaagaaattt gtttctcaaa ccgatgagge aaagaaatat catcagcacc	5280
	cattcttgcc actaacctai tcttacaage ctttaatect tcatttaagg cattagecaa	5340
	tccaacattt ctaggtaaag gtacaaatgt tactatttta ttgccaacat catcaatgaa	5400
	tgaatttata atatcgatgt gtgtttgatg gagttctcca tcgtcaacaa ttactatttg	5460
	atctggctta agtgtttgat cgtgaaaaat agagcgtaga gccacctcaa aaaattgagg	5520
	tagatcattt ttataaatgc taattaaaac tgagaatttt tctaacttat gattcatttc	5580
	attttaccac ttcgacccat taaaccgtca ttaatgcctt ttaaaaaaaaa atataacctt	5640
	tiattaccat ttggaaggaa aataggatat aaaaaaacct ttccaattaa ttaaccaga	5700
	ctagaaattt tccagtagat gggtaacata tttttattta ataaaagaaa gatatttcta	5760
	gtagcataat aaigacgaaa tgggtttggc aaaccgacag aaagaatatt taagatctta	5820
	aatgccecat ctccaagtct atgtgcaagt aacgcatttt tattctaat tactttaaac	5880
	ccagcagctc ttaatctcca acaatatcca tgggtctacc catcgataaa aagctcatct	5940
	ttcattctct caacaatcaa ccaactattt ttgggtatta gactgccaga acttaatgta	6000
	ctatctacct cataataaac ttctgtaagt ggtttccctt tttttacctt tgetttattt	6060
	aattaccag ttactttatc aaaatcttgt gaaccaacta aaccaacatt gacattttgt	6120
	ttaageaatt ttttgtaaca agtaagtaac tgcctiacca tcttaggata aggaatacta	6180

[0044]

```

tctgatcca ttgcaatai aaaatcagcg ccattttcaa aagcccatth cattctata 6240
ctttgggett ctgctatgcc taaattatca ttgaaattga atatttttac atcgctgaa 6300
gaattttcag catatttata accatttgta gagttattgc aaacgacaac tttagtaact 6360
tgtctcaaca ataattcaac cgcatthttt aaatcattat gttctgggtt gtaagcaacc 6420
aaaaaggcat atacagtgc catcttcacc ttaaaacctt catttagctt tcatcttttt 6480
tagaacatta cttaatgta ctaatacaat tattacagea acatggttag agtctaaaat 6540
ataaggatta gtaattgcat aagaacata tagaaaatat agcacacaca actcactgta 6600
ttttatgatt ttaatcgiga gaaggagatt aattaataaa aacaaagtaa ataaaaaac 6660
gccaagtga tttaaaaaat aaactgactg caattcataa tataatatatg cactataatc 6720
acggatagga gtttgaattt tgaigacatt acccaacca gaacctataa caaaatttga 6780
tacagactct gtaagatcat taattaatc agtaaaactga tcccatctaa ctctaaaga 6840
agaatcagct ccatttgatt tcatgattat caactcaatt gaatatgtaa taaaaaagg 6900
gagaatcaca gtaagaaaaa ccccaaaaat aattttctt aatttagctt atcgtgagtt 6960
agatttagaa catagtataa tatacataaa aaacaagcat atcgaaacaa aatatgcaaa 7020
attaccagcc actatagtae ctatagccag aataacggtt attgtatttt tgaatcgata 7080
atagaaaata tttttatga ctatagcaa cataaaggca aatggaatga gacatttcc 7140
tttaattga actctataga aaccacttcc atatgtataa acatcaccat aatcattctc 7200
caaaaaataa tgccttagtg ctgaataatc accaatacca tatgtttttg tcatataaat 7260
actaatgatg gatataataa ccgctgtaa taccattaaa tataaaaaa ttttaacaat 7320
cgagatggtt ccataagagc agaaataagc acataatata aataatatga taatataaaa 7380
cctaattatt atcgctatat cgttaccctt gatataggaa taaataaaat ttataaaaag 7440
agctaataga aatattaaaa taacaggata gtgatatttt ccgtttgcaa tttctttgt 7500
aatgacatg atacaaagac ataaaaacc ctccataatc caactatatt gaataaatgg 7560
aaagctacgt gtaaggaaaa atataaacc aaaaaacaa agaacactta aactttgtc 7620
ttttagatta taaaaatcag aagtcattgt tgcactctaa ttagatgggc ttgaggaagt 7680
aatccctaaa atcaattcgc tattaatatt tegtatcaat taataataat atcaaaaaat 7740
ctaacgatgt tcttacagac catgctattg cggtccaac aattcccaa tgataaataa 7800
aaatatataa taigcataaa tatgggataa ctccgagcaa atgaataata gctgtaattt 7860
ttgatcttcc actagcciga actgaaacaa atgggatttg tgcaatgcaa ttaaaaaaga 7920
aacctattgc aagaattttt aatactatac ctggcgctcc atgatatgta ggtcccatcc 7980
aagcggacat tataaaatct gataaaataa ttatcaacat tacaattgga agtataccaa 8040
taaccattat aaaaatgat aatattttag tttgctttac cgattgcaat tcigaactta 8100
atcttgaaa aatagctctg gacaacgcac ttggtaatat cgtaagcgt tgtatacctt 8160
cagacggagc agtataaaaa gaaactttat cageccccc aatgtgtgaa agaataaac 8220
gatccatata tgcataata gggctaaata tattgctaac tgttatccag ctccaaage 8280
cgattaatct tttaaetgtt acaattttta cagacagccc agatgatatt attagttttc 8340
gactaaatat aaaggtcact ataagtata agactcttgc cataactaaa ccatatatag 8400
cacttagtaa tctccaiga aaaaaacaga aaatcactgg taatccagcc acaaaagagt 8460

```

	tgtaaattga ttttattaaa tttacitttc tgaacttttc catcccccica aaaatcccca	8520
	accagacttg gttaacaag tataagggtg tggtagctga aataatatat attgcittga	8580
	cagattctac aacatgatic gegitaaigi llaataattl aacaattaca ttgctactca	8640
	aaaatagtag actaccgcca atcaagccca atatagttag aattaccgtt gaagttgaaa	8700
	tgatcgetct taattcitta tgaacatttt tatatattga tacttetett ataacagete	8760
	tggfcaatcc agcatcaaaa ataactigcat atccaactaa ggeaatagct aacgtaaaaa	8820
	ggccaaattg ctgggtccct agaattctag acagtatacc taacgcagga attgcattta	8880
	atgatgggat aatataccca ettatatcc ataaagtatt ctttacaata ctcacaaaaa	8940
	taattccctc atgttatgca attcttttagc ccttgcactc ttaatcgata aaatataatt	9000
	attatgttct atcgtcgccc attttatget cagaatagga tcattccata caateccctc	9060
	atcactatca ggaatataat agttcgtcgt tttatataaa aattccgcag tctcgtcag	9120
	cacccaaaaa ccattgtgcaa atccctcagg gatecacaat tgcgccttat tctcagcaga	9180
	taaatccace ccaacccatt taccaaaggt aggcgacgat ttacgaatat caacagctac	9240
	atcaaaaacc tcaccaacaa cgcaacgtac cagtttccct tgcgcataag gttctaaactg	9300
	ataatgcage ccgcgtaaaa caccittact agacttcgaa tggttatcct gaacaaattc	9360
	aaccittact cctacagetti cttcgaaaaa tttctgatta aagctttcca taaagaaacc	9420
	acgtcatca ccaaaaaact tggctcgaa aattaacaca tcaggaattt ctgttttaatt	9480
	tacgttcatt ttatttaataa cttttaatca ttttcagcag atactgtcca taagcatitt	9540
[0045]	ttttcagcgg ctccgctaant gtttcacct gtccagcacc aataaacctt ttacggtaag	9600
	caatttcttc tgggcaggaa accttiagtc cctggcgctc ttcaatgggtg gcaatgaagt	9660
	tgtttgttcc aataagactc tgaigtgtcc ccgtatccag ccattgcataa ccacgcccc	9720
	tcattgcaac ggataaacgc cctgtttcca tataataacg gttaatatcg gtaatttcca	9780
	gttcaccacg ggcagaagge ttaaggtttt tgcceatttc gacaacgtcg ttatcataga	9840
	aataaagccc ggttaccgca taattacttt ttggtttagt cggtttttct tccaggctta	9900
	ttgcctgacc gtttttatca aaactaacga cgccgtagcg ttcaggatca ttaacgtgat	9960
	aggcaaatat cgttgcacca cttcttttgt taacagcgac atccattaac ttcggcagat	10020
	catgaccgta gaagatatta tcaccaagaa ccaaagcaca atcaccacca ccgataaact	10080
	cttcaccgat aataaacgcc tgcgaagcc catctggagt cggttgcact ttgtaactga	10140
	gatttagccc ccactggcta ccgtcaccta gcagttgttg aaaacgagga gtatcctgtg	10200
	gcgtactaat aatcagaata tcgcgaatcc ccgccaacat cagtgtagag agcgggtaat	10260
	agatcatcgg cttatcataa ataggtaata gctgtttact gacagccata gtcacaggat	10320
	aaagacgtgt accagaacca cccgctaaaa taataccttt acgcgttttc atttcacat	10380
	tccttttaatt tcattcttgt ccaccatcac gaacaagatg caaaaactat taaattgtg	10440
	tagtctgaat taattcgttg agcattcgtt tcacaccaac ctgccagtea ggcaagacaa	10500
	gcgcaagttt ctgctgaaat tttctgtiat taaggcgaga gttatgtgga cgacgagctg	10560
	gtgtaggata ggctgttgtt ggtacitcgt ttagcttgtt gactgcaagg ggaatacctg	10620
	ctttgcgcgc ctcttcaaaa accagcgcag cataatcgtg ccaggttgtg gtaccactgg	10680
	ctaccagatg gtacaaacct gcgacttccg gtttattcag tgccacacga atagcatgtg	10740

[0046]

```

ccgtacaatc agccagcagc tcagcacctg ttggegcacc aaattgatca tttatcacag 10800
ccagttcttc gegetctttt gccagacgca acatcgtttt ggagaagtta tttccttag 10860
ctgcgtatac ccagctggta cggaaaataa gatgcttcgc gcaatgttcc tgtaacgctt 10920
tttctccggc taacttgggt tcaccgtaaa catttagcgg tcgggttgea tcegtctcca 10980
gccatggcgt gtcgccattt ccagggaata cgtagtcagt tgagtaatga attaccecaag 11040
ccccaactic attagcctct tttgcaattg attcaacaet agtcgcattg agtaattgtg 11100
caaattcggg ttctgactca gccattatcta ctgcgggtgtg agccgcagca ttaacaataa 11160
catcaggctg aattcttttg actgtttcag ctacacctic aggattacta aaatcaccac 11220
aalaatcagt ggagtgaaca tcaagagcaa lcaaatlace caaaggigcc agagcacgct 11280
gtagtccca acciacctgc cctgttttgc cgaaaaggag gatattcatt actggegccc 11340
ctcatagttc tgttcaatcc acgattgata agcaccacit ttcacattat caacccattt 11400
tgtattggac aggtaccatt ccaatgtctt ccgaatcccg ctctcaaacg tttcctgcgg 11460
tttcagccc aattcgcggc taactttctc tgcatcaatc gcataacggc gatcgtgtcc 11520
cggcgatcg gcaacataag taatttgcic gcggtaagat ttctctttcg gtacaatctc 11580
atccagcaaa tcacaaatag tgagcactac atcgatgttt ttcttttctg tgtgtccacc 11640
aatgttataa gtttcacccc ctttaccctic ggttacgacg gtatataacg cagcgcatg 11700
atcttcaaca tacagccagt caggaatttg atcccccttg ccataaatag gtaatgcctt 11760
accttcaga gcattcagaa taaccaatgg aatcaatttt tccgggaaat gataaggacc 11820
ataattatta gagcaattag tcacaatggt tggtaacea taggtacggt tccacgcgcg 11880
gactaaatga tcgctggatg ctttgaagc ggaataaggg ctgcttggcg cgtaagcigt 11940
tgtctctgta aataaggcta attcttctgt attatttacc tegtccaggat gaggcacatc 12000
accatagact tcgtcagtag aaatatgatg aaaaacggaat ctagtcttct tgcctctatc 12060
aagagcagac caataattgc gacgcgcttc caaaaggaca tatgtacca caatatgggt 12120
ttcaataaat gccgcaggac ctgtaattga acggtcaaca ttgctttcag cagccagggtg 12180
catactgca tctggctgat gctgagcaaa aatccgtgcc attgcagctg categcaaat 12240
atccgatgt tcaaaaacat agcgttcaga atcagaaaca tcagcaagtg attccagggt 12300
tcggcgctac gtttaatttat cgacattaac aacactatcc tgcgtattat ttataatgtg 12360
acgaactaca gcaaaaccaa taaatccigc gccaccagta acaagtattt tcacctaat 12420
tattccatat tcttcagag catgeigtga aataagcgge tctcagtttg attaatagaa 12480
gtattaatgc acgtaccgc ccttggtttt acagctacca gagcactgca tgeatgecta 12540
cgatgtgacg agcgttaccc actcgcgcta aacccgaaaa attcaaaagc taattgtctt 12600
accaatccgc tctggaacaa aggaaaatcc tggaaaactt tgactaaaat cctattgcta 12660
actcgttggt attctgattg tttatataaa acaacggcag gaattatcgc aacaaattac 12720
tttcaccacg aatcttctact gccgttataa ttttcttate aaccgttaca tccggtcaga 12780
tttctattat tctgttaaca gttctcaat acctttacgg aacttgcgcc cttctttcag 12840
gttgcgcage ccatacttca caaacgcctg catatagccc atttttttac cgcagtcgta 12900
gctgtgcceg gtcacagca ttgcatcaac ggaactgttt ttgcgcagct cggcaatggc 12960
atcagtcage tgaatacgtc cccatgcacc aggetgagta cgttcaagtt ccggccaaat 13020

```

	atcggcagaa agcacatagc gaccaacggc caigatgtct gagtccagcg tcigcggtg	13080
	atccggtttt tcgataaatt caacaatgcg gctgacttta cettccgat ccagcggttc	13140
	tttggctcgg atgacggagt attcagagag gtcacccggc atacgttttg ccagcaccig	13200
	gtacggccc gtttcaitga agcgcgcaat catggcagca aggttgtagc gtagcgggtc	13260
	ggcgctggcg tcgtcgatca caacgtctgg cagcaccacg acaaatggat tgcaccaat	13320
	ggcggttcgt gcacacaaaa tggagtgacc faaacctaaa ggttcgccc tgcgcacgtt	13380
	cataatagtc acgccccggc ggcagataga ttgcacttcc gccagtagtt gacgcttcac	13440
	gcgtgtctca aggagagatt ctaattcata agaggtgtcg aagtgtttt cgaccggtt	13500
	cttggacgca tgagttacca ggaggatttc ttgatccct gcagccacaa tctctcaac	13560
	aatgtactga atcattggtt tgtegacgat cggtagcaco tctttgggta tgccttagt	13620
	ggcaggcaac atatgcaccc caagaccgc taccggtata actgccttta aattcgtcat	13680
	tattttecta cctctaaggg gctgatagtg cgtaaattat tgctataggt tagccaaacg	13740
	gtatggctat ataccaagca taactttgat faaaccttac gataacacta cacaccata	13800
	gcactcgggt tactcggatt actcggaaat ccacatactg ataatttaac cagtacctct	13860
	ttccgaataa tcgtagtcga acctggctct ttttctctg actcgtctgc attactcaga	13920
	aacaaacgtt atgtcgtctt ttttggcatg gacgaattca tactgcagag ttgatccag	13980
	accttgcgac agegtatacg gtgcacaaa acctgaagaa tgcactttcg ttgcgtcaaa	14040
	ctgtgttgtt gcgcagaatt tttcacgcg cacagagctg acagcgtatt tttgcccgt	14100
[0047]	aattttgctc aggatacna agcaatacc acccagcatt cctagfgggt aaggcaagtg	14160
	catagaaggg atctttttgt tcaggttttg ttcaacttca gcaaccaact ggttcagtgt	14220
	caggctcggc ttatcaacat agttataaac ctcataacct gcggcaacat tcttcagttt	14280
	gtacttgata aactcaacaa tgtttccaac ataagccatg gacttatagt tagtccctgc	14340
	gcccaccatc ataaacttgc cgcagcgat ctgtttcagc aagttataga cgttaccgcg	14400
	gttgcgttca ccgaagataa cggtaggacg gatgatggtt aatgaacgtt ctgttgggtc	14460
	ttgtttatac cattcagca gcaettctct tgccgtccac ttacttttgc cgtagtgtt	14520
	gaaagggtcg tgtggatggt ttctgtcagg gtigtgtttg ttcaaaccat aaacagcaac	14580
	ggaactggta aagatgatat ttttaacgcc atttttttcc atggccgcca gcacattgcg	14640
	ggtaccctga acgttgacat cataatagag agaagtaggg ctgacgtcat cgcggtggtc	14700
	cgtctgccagt agtacaacag tgcacaaacc ggctaacgcc tggtcgagtg cctgttgatc	14760
	acgaacatca ccaatctgtg tgatttctgg ataaaagtgg cctctgccgtt tgtccaggtt	14820
	cttgatatta aagtcagcaa ttgccgtttc aagtagtcgg gtctctacga atccggaagc	14880
	tctatgagc aaaaagtat tgttcataaa tcactttagt ctggttggtta cgtagaacac	14940
	acaagataaa gatgagtacc ttccctgagt agtcaatgct gccagcccc agctttaaca	15000
	gttagtgtga ggattataat cttttagaac attatataca gtaagtatat gaatggctgc	15060
	aatctactc tctccgttcc ggcaatctaa agttaatgct agcgacgtcg tgggactctc	15120
	tagagtcgac ctgcaggcat gcaagcttga gtattctata gtctcaccta aatagcttgg	15180
	cgtaatcatg gtcatactg tttctgtgt gaaattgtta tccgtcaca attccacaca	15240
	acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaaacta	15300

	cattaatigc gttgcgctca ctgcccgtt tccagtcggg aaacctgtcg tgccagctgc	15360
	attaatgaat cggccaacgc gaaccccttg cggccgcccg ggccgtcgac caattctcat	15420
	gtttgacage ttatcatcga atttctgcca ttcatccgct tattatcaact tattcaggcg	15480
	tagcaaccag gcgtttaagg gcaccaataa ctgccttaaa aaaattacgc cccgccctgc	15540
	cactcatcgc agtactgttg taattcatta agcattctgc cgacatggaa gccateacaa	15600
	acggcatgat gaacctgaat cgcacgccc atcagcacct tgctgccttg cgtataatat	15660
	ttgcccctag tgaatacggg ggcaagaag ttgtccatat tggccacgtt taaatcaaaa	15720
	ctggtgaaac tcaccaggg attggtgag acgaaaaaca tattctcaat aaaccttta	15780
	gggaaatagg ccaggttttc accgtaacac gccacatctt gcgaatatai gtgtagaaac	15840
	tgccggaaat cgtcgttgta ttactccag agcgatgaaa acgtttcagt ttgctcatgg	15900
	aaaacggtgt aacaagggtg aacactatcc catatcacca gctcacctgc ttctattgcc	15960
	atacgaatct cggatgagc attcatcagg cgggcaagaa tgtgaaataa ggccggataa	16020
	aacttgtgtt tatttttctt tacggtcttt aaaaaggccg taatatccag ctgaacggtc	16080
	tggttatagg tacattgagc aacfgactga aatgctcaa aatgttcttt acgatgccat	16140
	tgggatatai caacggtgtt ataccagtg atttttttct ccatitttagc ttccittagct	16200
	cctgaaaate tcgataacte aaaaaatacg cccggttagtg atcttatttc attatggtga	16260
	aagttggaac ctcttacgtg ccgatcaacg tctcatcttc gccaaaagti ggcccagggc	16320
	ttcccgttat caacagggac accaggattt atttattctg cgaagtgate ttccgtcaca	16380
[0048]	ggtatttatt cgcgataagc tcatggagcg gcgtaacegt cgcacaggaa ggacagagaa	16440
	agcgcggatc tgggaagtga cggacagaac ggtcaggacc tggattgggg aggcggttgc	16500
	cgccgctgct gctgacggtg tgacgttctc tgttccggtc acaccacata cgttccgcca	16560
	ttcctatgcg atgcacatgc tcatgcccgg tataccgctg aaagttctgc aaagcctgat	16620
	gggacataag tccatcagti caacggaagt ctacacgaag gtttttgcgc tggatgtggc	16680
	tgcccggcac cgggtgcagi ttgcgatgcc ggagctgat cgggttgcca tgcgaaaca	16740
	attaacctga gaataaatgc cttggccttt atatggaaat gtggaactga gtggatatgc	16800
	tgtttttgtc tgttaaacag agaagctggc tgttatccac tgagaagega acgaacagti	16860
	cgggaaaate tccatttate gttagatcc gcattattaa tctcaggagc ctgtgtagcg	16920
	tttataggaa gtatgtttct gtcatgatgc ctgcaaggcg taacgaaaac gatttgaata	16980
	tgccttcagg aacaatagaa atcttctgtc ggtgttacgt tgaagtggag cggattatgt	17040
	cagcaatgga cagaacaacc taatgaacac agaaccatga tglggctctgt ccttttacag	17100
	ccagtatgct tgcgcgcatg cgcgcgacag ggccgaagccc tgggttggtt gccctcgccg	17160
	ctgggctggc ggccgtctat ggccctgcaa acgcgccaga aacgcgctcg aagccgtgtg	17220
	cgcacacccg cggccggccg ccggcgttgt ggatacctcg cggaaaactt ggccctcact	17280
	gacagatgag gggcggacgt tgacacttga ggggcgcact caccgcgcgc ggcttgaca	17340
	gatgaggggc aggtctgatt tgcgcggcg acgtggagct ggccagcctc gcaaatcggc	17400
	gaaaacgctt gattttacgc gagtttccca cagatgatgt ggacaagctt ggggataagt	17460
	gccctcggtt attgacactt gagggcgcg actactgaca gatgaggggc gcgatecttg	17520
	acacttgagg ggcagatgct tgacagatga ggggcgcacc tattgacatt tgaggggctg	17580

	tccacaggca gaaaatccag catttgcag ggtttccgce cgtttttcgg ccacccgtaa	17640
	ccgtctcttt aacctgcttt taaaccaata ttatataacc ttgtttttta ccagggcctgc	17700
	gccctgtgcg cgtgaccgag cagcccgag gggggtgcce ccccttctcg aacctcccg	17760
	gtcagagtga cgaggaagca ccagggaaca gcacttata atctctgcta cacacgatgc	17820
	ctgaaaaaac ttcccttggg gttatecact tatecacggg gatattttta taattatttt	17880
	ttttatagtt tttagatctt cttttttaga gcgccttgta ggccctttatc catgcttggt	17940
	ctagagaagg tgttgtgaca aattgccctt tcagtgtgac aaatcacctt caaatgacag	18000
	tcctgtctgt gacaaatgc ccttaacctt gtgacaaatt gccctcagaa gaagctgttt	18060
	tttcacaaag ttatccctgc ttattgactc ttttttatit agtgtgacaa tctaaaaact	18120
	tgfcacactt cacatggatc tgtcatggcg gaaacagcgg ttatcaatca caagaaacgt	18180
	aaaaatagcc cgcgaatcgt ccagtcaaac gacctcactg aggcggcata tagtctctcc	18240
	cgggatcaaa aacgtatgct gtatctgttc gttgaccaga tcaganaaat tgatggcacc	18300
	ctacaggaaac atgacggtat ctgcgagatc catgttgtta aatatgctga aatattcgga	18360
	ttgacctctg cgggaagccag taaggatata cggcaggcat tgaagagttt cgcggggaag	18420
	gaagtgggtt ttatcgcce tgaagaggat gccggcgatg aaaaaggeta tgaatctttt	18480
	cccttggttta tcaaacgtgc gcacagtcca tccagagggc ttacagtggt acatatcaac	18540
	ccatatctca ttcccttctt tategggtta cagaaccggt ttacgcagtt tggcttagt	18600
	gaaacaaaag aaatcaccaa tccgtatgcc atgcgtttat acgaatccct gtgtcagtat	18660
[0049]	cgtaagccgg atggctcagg catcgtctct ctgaaaatcg aciggatcat agagcgttac	18720
	cagctgcctc aaagttacca gcgtatgctt gacttcgccc gcccttccct gcaggctctgt	18780
	gttaatgaga tcaacagcag aactccaatg cgcctctcat acattgagaa aaagaaaggc	18840
	cgccagacga ctcatatcgt atttccctc cgcgatata ctcccatgac gacaggatag	18900
	tcgtagggtt atctgtcaca gatttgaggg tggttctgta catttgctct gacctactga	18960
	gggtaatttg tcacagtttt gctgtttcct tcagcccgca tggattttct catacttttt	19020
	gaactgtaat ttttaaggaa gccaaatttg agggcagttt gtcacagttg atttccctct	19080
	ctttcccttc gtcatgtgac ctgatatcgg gggttagttc gtcacattg atgagggttg	19140
	attatcacag ttattactc tgaattggct atcccgctgt gtacctctac ctggagtttt	19200
	teccacggtg gatatttctt cttgcctga gcgtaagagc tatctgacag aacagttctt	19260
	ctttgcttcc tcgccagttc gctcgtatg ctccgttaca cggctgcgga gagecctagt	19320
	gataataagt gacigaggta tgtgtctctt ttatctcctt ttgtagtgtt gctcttattt	19380
	taaaacactt tgcgggtttt tgatgaattt gcgattttgt tgttgccttg cagtaaatg	19440
	caagattttt taaaaaacg caaagcaatg attaaaggat gttcagaatg aaactcatgg	19500
	aaacacttaa ccagtgcata aacgttggtc atgaaatgac gaagctatc gccattgcac	19560
	agtttaatga tgacagcccg gaagcgagga aaataaccgg gcgctggaga ataggtgaag	19620
	cagcggattt agttggggtt tcttctcagg ctatcagaga tgcggagaaa gcagggcgac	19680
	taccgcacce ggatattgaa attcgaggac ggggttgagca acgtgttggt tatacaattg	19740
	aacaaattaa tcatatgcgt gatgtgtttg gtacgcgatt gcgacgtgt gaagacgtat	19800
	ttccacgggt gatcggggtt gctgcccata aaggtagcgt ttacaaaacc tcagttcttg	19860

	ttcatcttgc tcaggatctg gctctgaagg ggctacgtgt ttgtctctg gaaggtaacg	19920
	acccccaggg aacagcctca atgtatcacg gatgggtacc agatcttcat attcatgcag	19980
	aagacactct cctgcccttc tatcttgggg aaaaggacga tgtcacttat gcaataaagc	20040
	ccacttctg gccggggctt gacattatc cttectgtct ggctctgcac cgtattgaaa	20100
	ctgagttaat gggcaaattt gatgaaggta aactgccac cgatccacac ctgatctcc	20160
	gactggccat tgaaacigtg gctcatgact atgatgtcat agttattgac agcgcgccta	20220
	acctgggtat cggcacgatt aatgtcgtat gtgtctctga tgtctctgatt gtccccagc	20280
	ctgtctgagt gtttgaetac acctccgcac tgcagttttt cgatatgctt cgtgatctgc	20340
	tcaagaacgt tgatcttaaa gggttcgagc ctgatgtacg tattttgctt accaaataca	20400
	gcaatagtaa tggctctcag tccccgtgga tggaggagca aattcgggat gcctggggaa	20460
	gcatggttct aaaaaatggt gtacgtgaaa cggatgaagt tggtaaaggt cagatccgga	20520
	tgagaactgt ttttgaacag gccatigatc aacgctcttc aactgggtgc tggagaaatg	20580
	ctcttctat ttgggaacct gtctgcaatg aaattttcga tctgtctgatt aaaccacgt	20640
	gggagattag ataatagaag gtgcgccigt tattccaaaa catacgtcct atactcaacc	20700
	ggttgaagat acttcgttat cgacaccagc tgcctccgat gtggattcgt taattgcgcg	20760
	cgtaggagta atggtctcgc gtaatgccat tactttgcct gtatgtggtc gggatgtgaa	20820
	gtttactctt gaagtgtccc ggggtgatag igttagaag acctctcggg tatgttcagg	20880
	taatgaacgt gaccaggagc tcttactga ggacgcactg gatgatctca tcccttcttt	20940
[0050]	tctactgact ggtcaacaga caccggcggt cggctgaaga gtatctgggt tcatagaaat	21000
	tgcctgatgg agtcgccgtc gtaaagetgc tgcacttacc gaaagtgatt atcgtgttct	21060
	ggttggcgag ctggatgatg agcagatggc tgcattatcc agattgggta acgattatcg	21120
	cccaacaagt gcttatgaac gttgtcagcg ttatgcaagc cgattgcaga atgaatttgc	21180
	tggaaatatt tctgcctgg ctgatcgga aaatatttca cgtaagatta ttaccgcgtg	21240
	tatcaacacc gccaaattgc cttaateagt tgttgccttt ttttctcacc ccggtgaact	21300
	atctgcccg tcagggtgat cacttcaaaa agcctttaca gataaagagg aattacttaa	21360
	gcagcagga tctaaccttc atgagcagaa aaaagctggg gtgatatttg aagctgaaga	21420
	agttatcact cttttaactt ctgtctttaa aacgtcatct gcatcaagaa ctagttaaag	21480
	ctcacgacat cagtttgcct ctggagcgac agtattgtat aagggcgata aaatggtgct	21540
	taacctggac aggtctctgt tccaactga gtgtatagag aaaattgagg ccattcttaa	21600
	ggaacttgaa aagccagcac cctgatcgga ccacgtttta gtctacgttt atctgtcttt	21660
	acttaattgc ctttgttaca ggccagaaag cataactgac ctgaatatte tctctgggac	21720
	cactgttcca ctgtatctg cgtctgata atcagactgg gaccaaggte ccactcgtat	21780
	cgtcggctct attattagtc tgggaccaag gtcccactcg tategtcgtt ctgattatta	21840
	gtctgggacc aaggteccac tctatctg ggtctgataa tcagactggg accaaggtec	21900
	cactcgtatc gteggctcga ttattagctt gggaccatgg tcccactcgt atcgtcggtc	21960
	tgattattag tctgggacca cgttccactt cgtatctctg gtctgattat tagtctggaa	22020
	ccacggtccc actcgtatcg tgggtctgat tattagctct ggaccacggt cccactcgt	22080
	tctcggctct gattattagt ctgggaccac gateccactc gtgttctcgg tctgattate	22140

```

ggctctgggac caeggteccca cttgtattgt cgatecagact atcagegtga gactacgatt 22200
ccatcaatgc ctgtcaaggg caagtattga catgtctctg taacctgtag aacggagtaa 22260
cctcggtgtg cggttgtatg cctgtctgtg attgtctctg tgctctgctt atccacaaca 22320
ttttgcgcac ggttatgtgg acaaaatacc tggttacca gcccggtgccc gcacgttaac 22380
cgggctgcat ccgatgcaag tgtgtcctg tgcacgagct cgcgagctcg gacatgaggt 22440
tgccecgat tcaagtctgc tgatttgtat tgtctgaagt tgtttttacg ttaagttgat 22500
gcagatcaat taatacgata cctgcgtcat aattgattat ttgacgtggt ttgatggcct 22560
ccacgcacgt tgtgatatgt agatgataat cattatcaact ttacgggtcc ttccggtga 22620
tccgacaggt tacggggcgg cgacctcgcg ggttttcgct atttatgaaa attttccggt 22680
ttaaggcggt tccgtttctt ttcgtcataa cttaatgttt ttatttaaaa taccctctga 22740
[0051] aaagaaagga aacgacaggt gctgaaagcg agctttttgg cctctgtcgt ttcctttctc 22800
tgtttttctc cgtggaatga acaatggaag tccgagctca tcgctaataa ctctgtatag 22860
catacattat acgaagttat attcgat 22887

```

<210> 30

<211> 6

<212> PRT

<213> 人工序列

<220>

<223> 人工序列描述: 合成
6xHis标签

<400> 30

His His His His His His

1 5

=

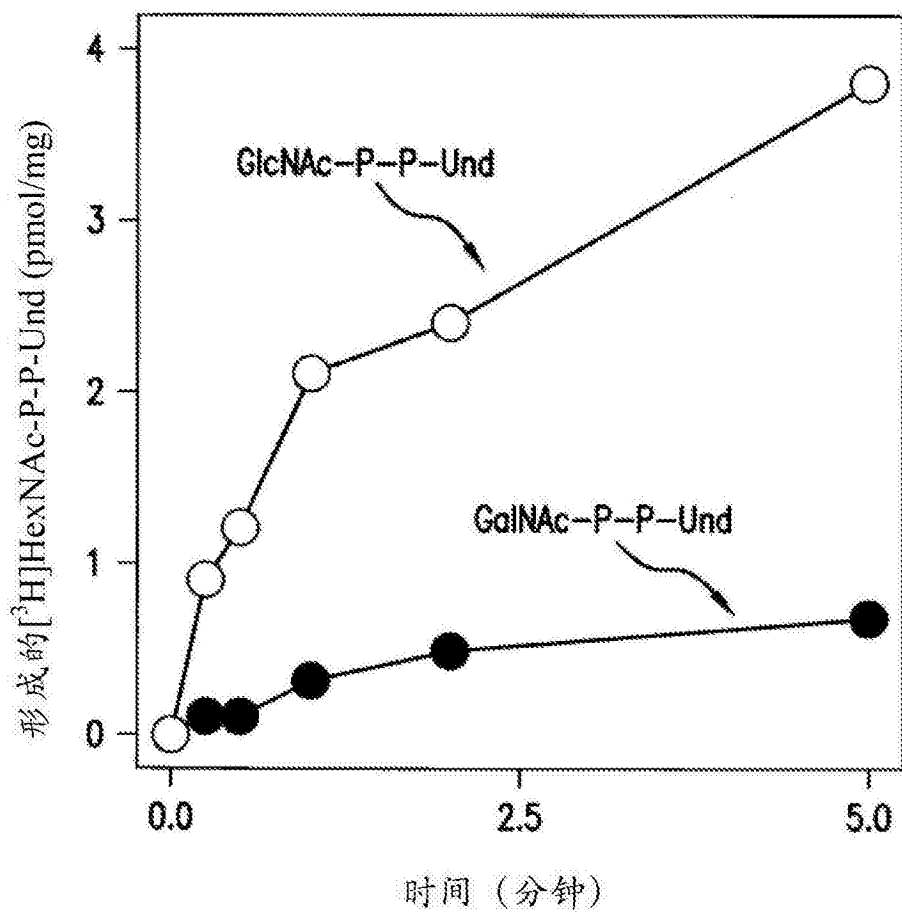


图1

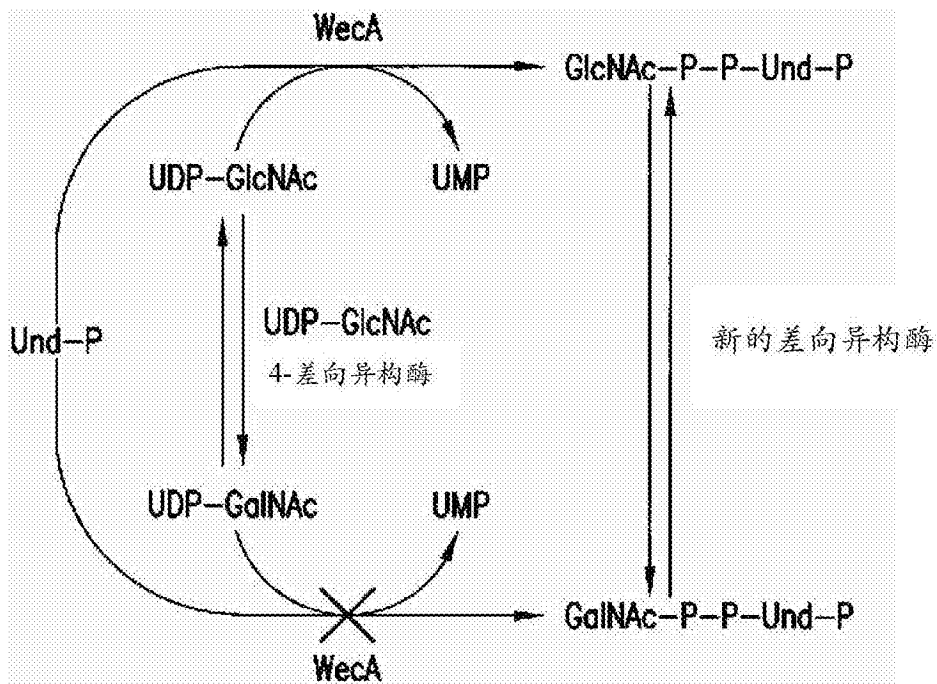


图2

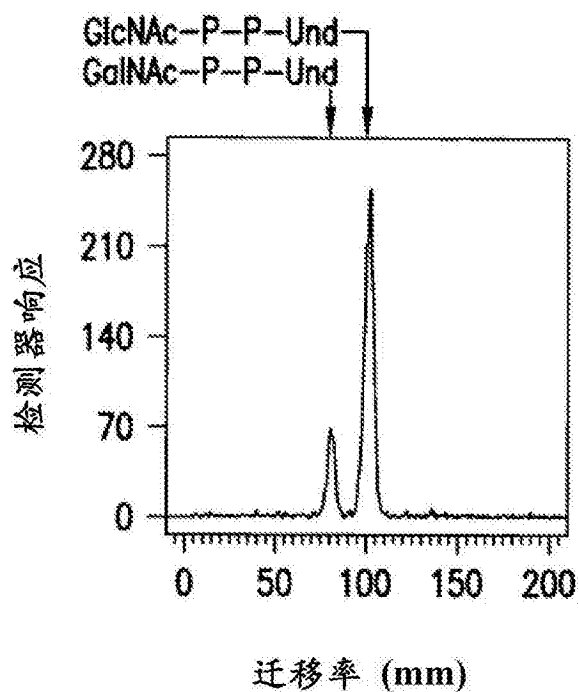


图3A

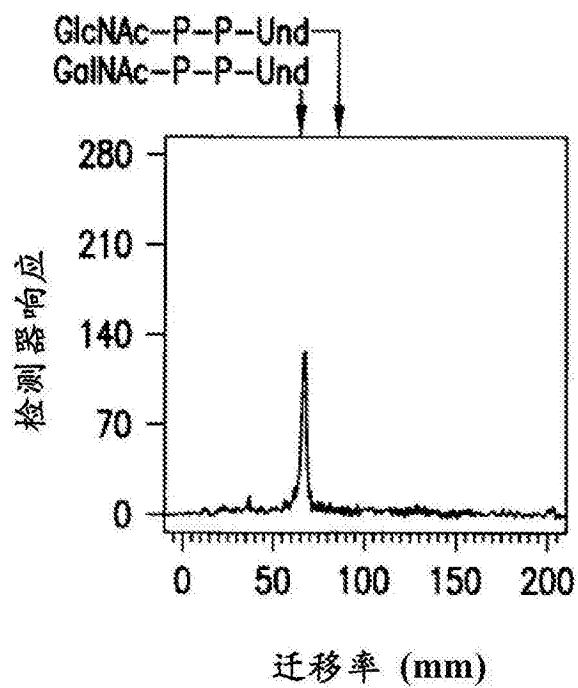


图3B

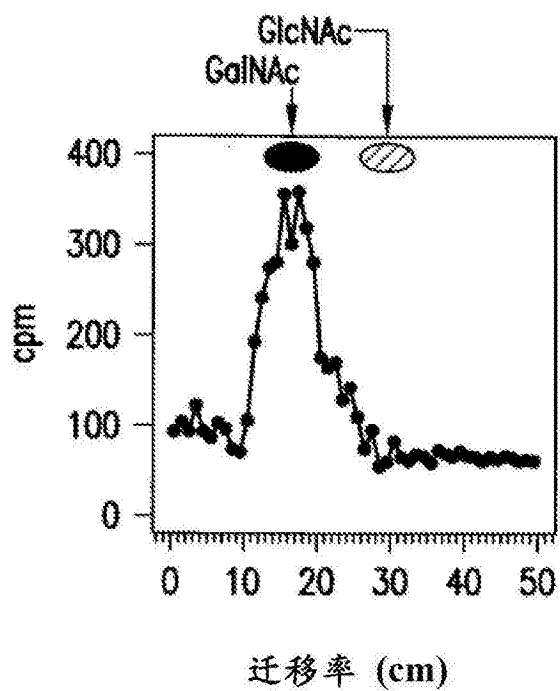


图3C

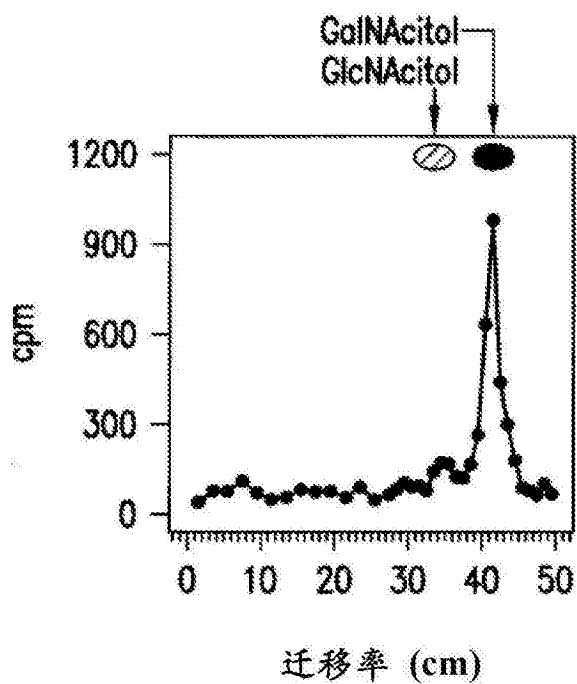


图3D

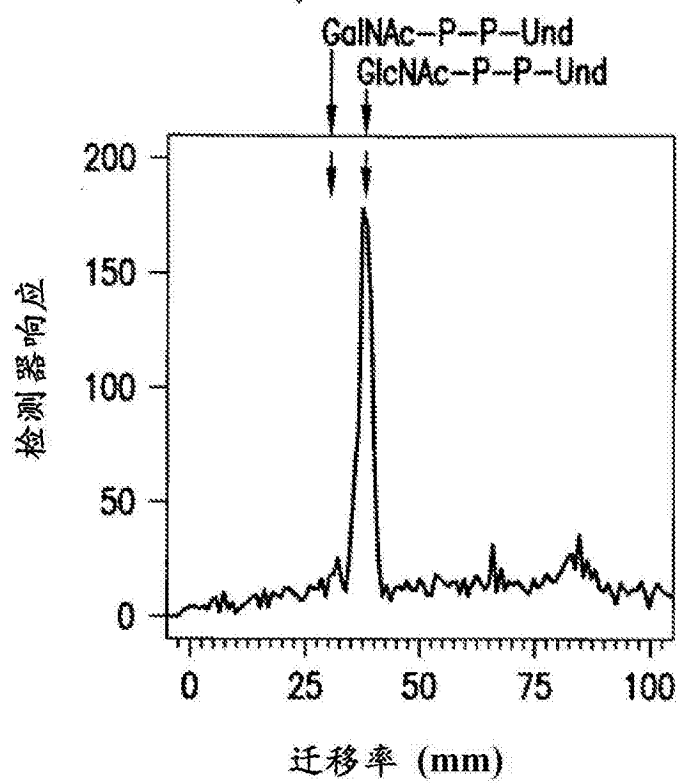


图4A

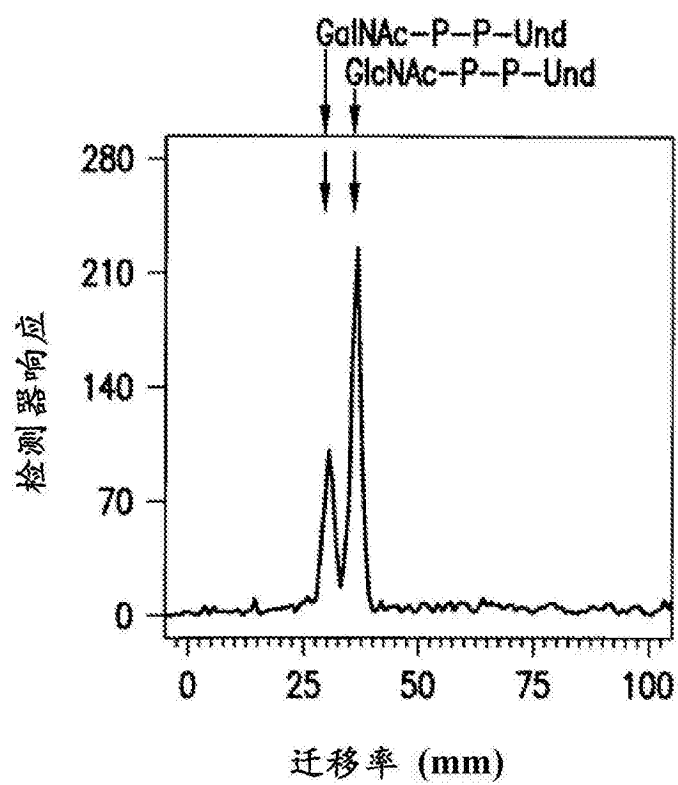


图4B

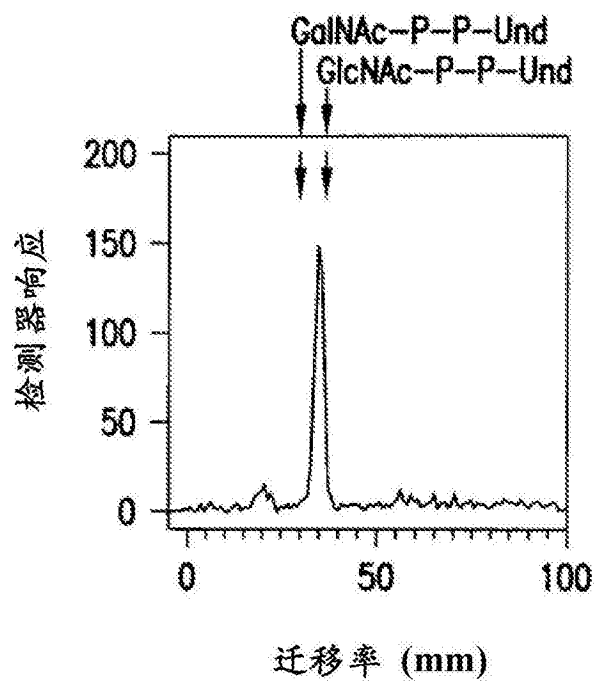


图5A

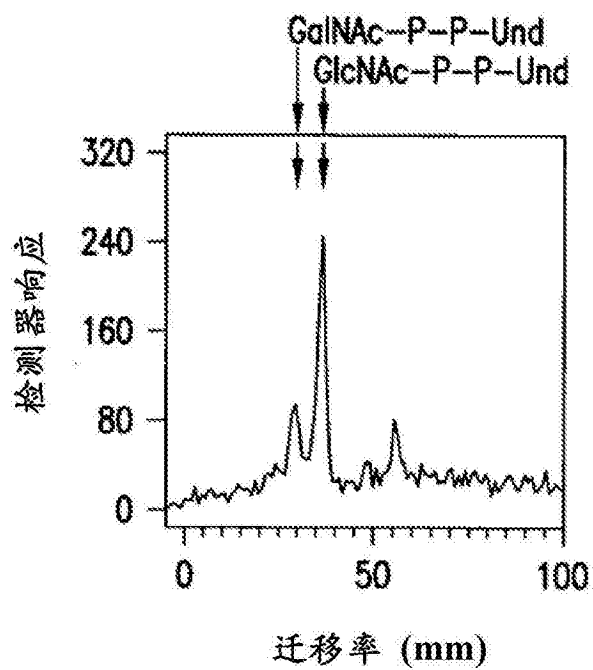


图5B

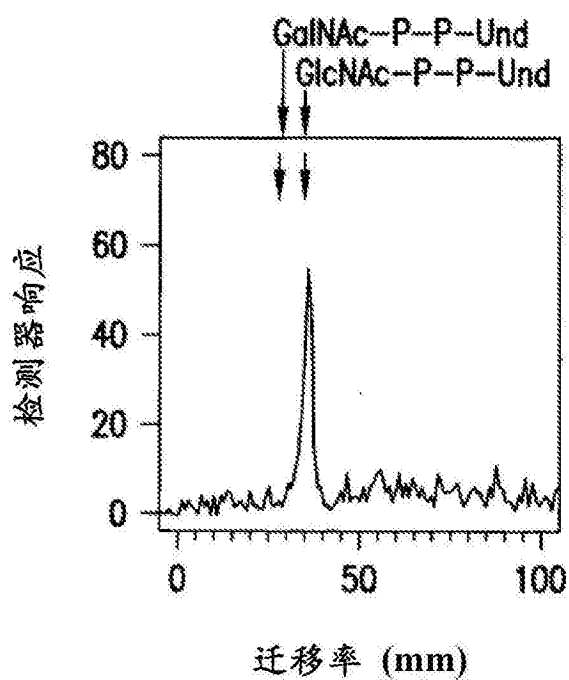


图5C

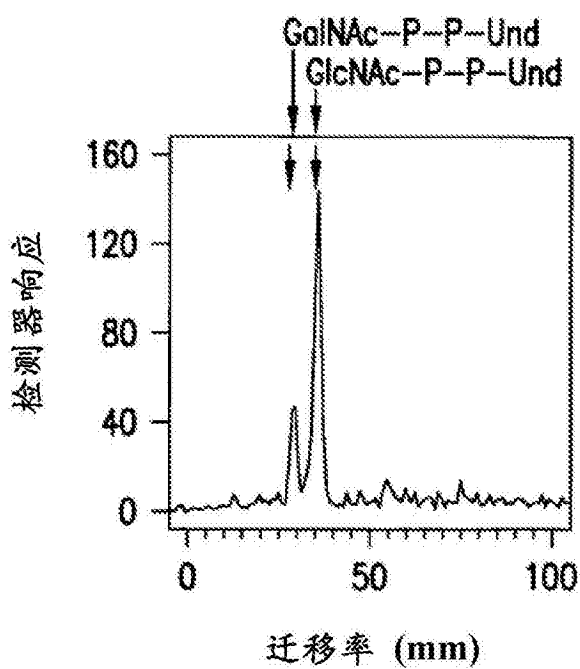


图5D

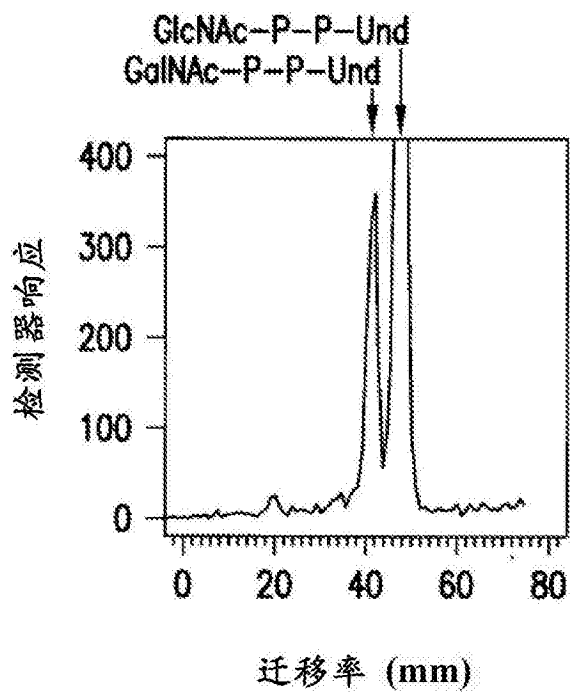


图6A

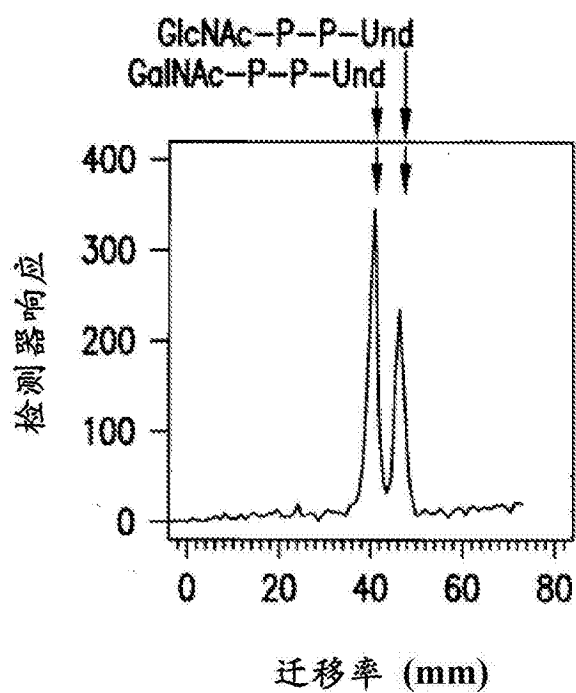


图6B

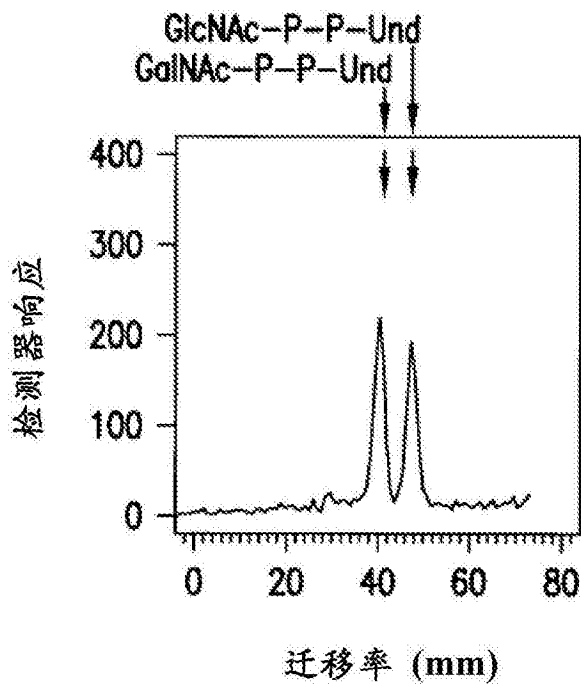


图6C

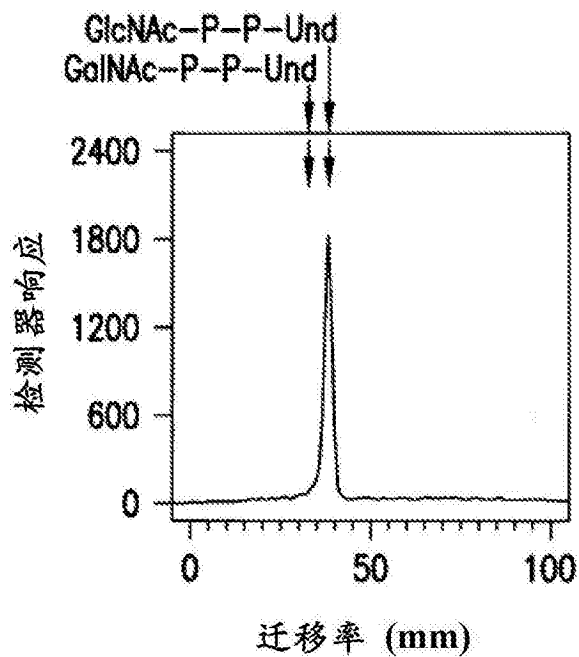


图7A

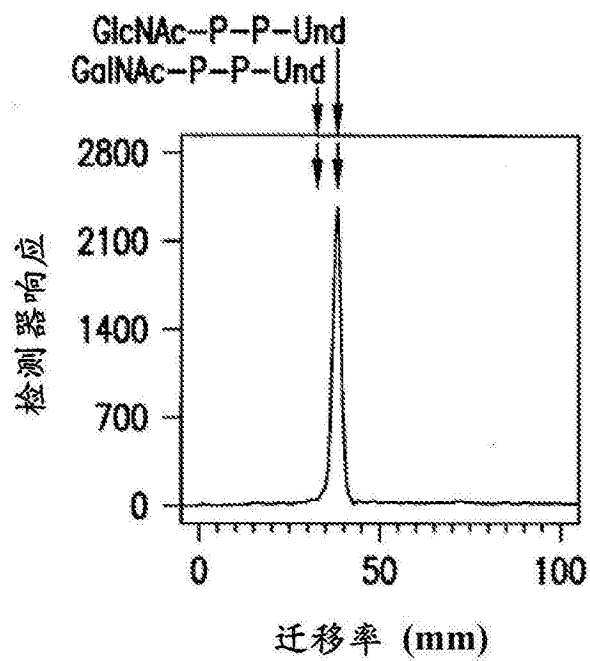


图7B

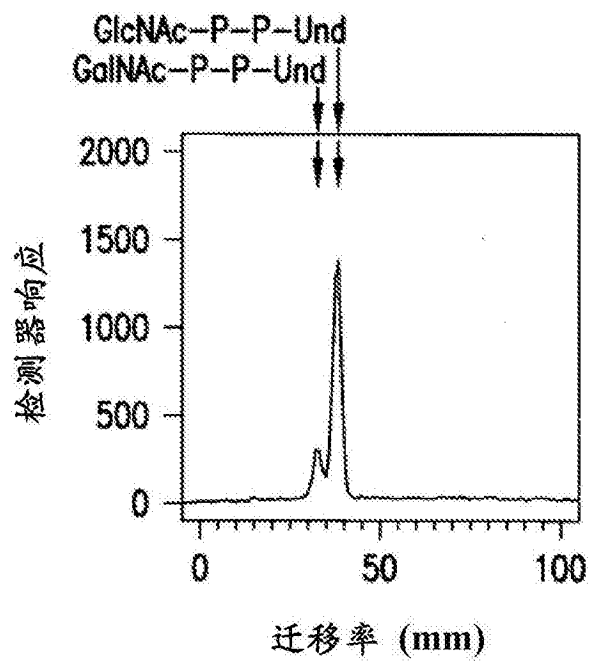


图7C

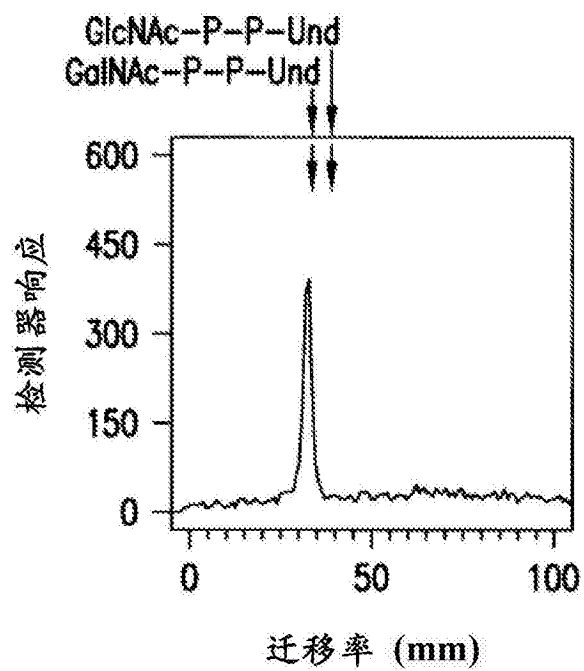


图7D

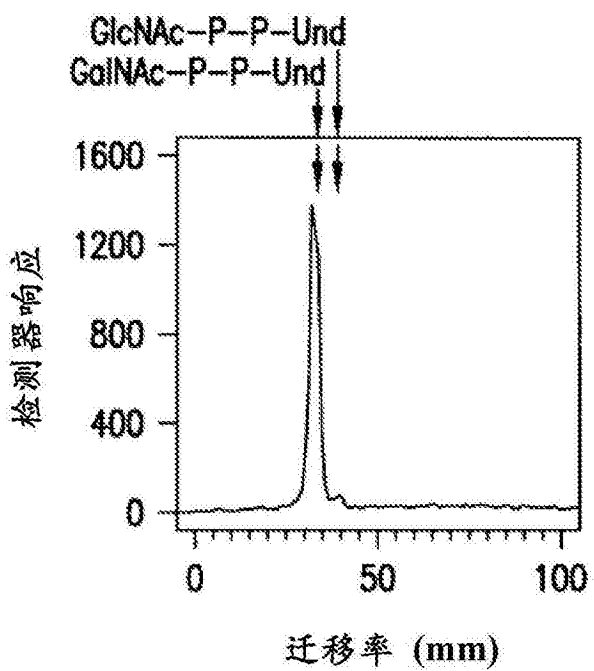


图7E

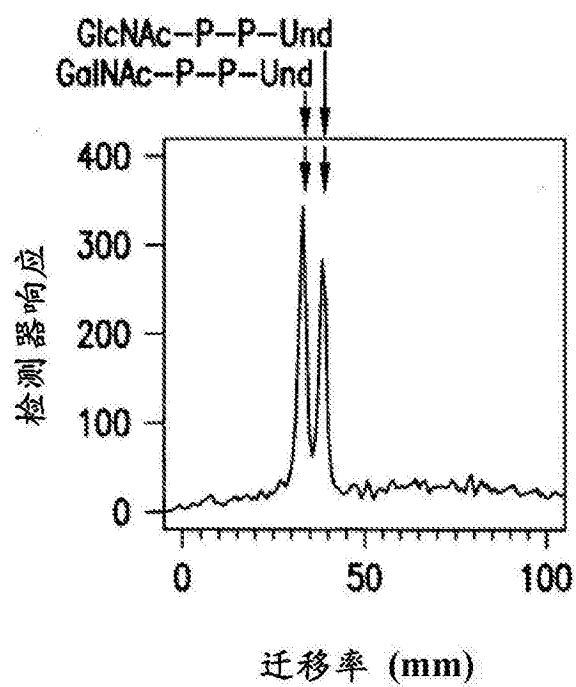


图7F

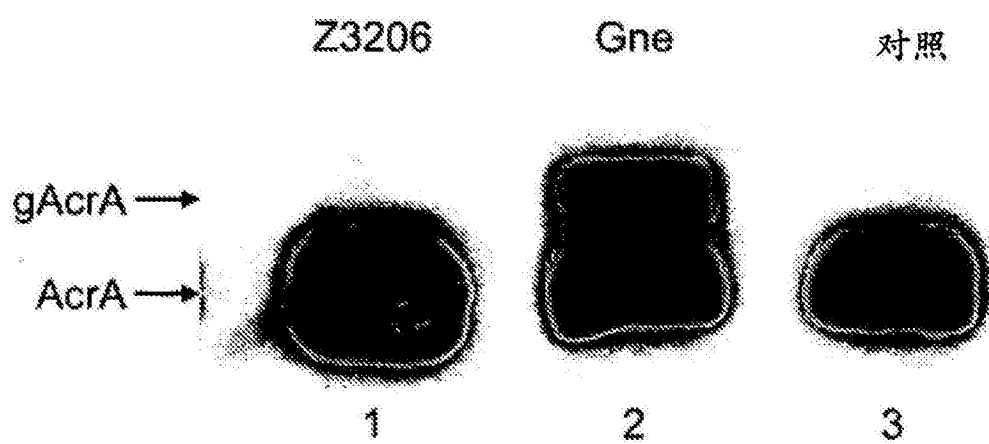


图8

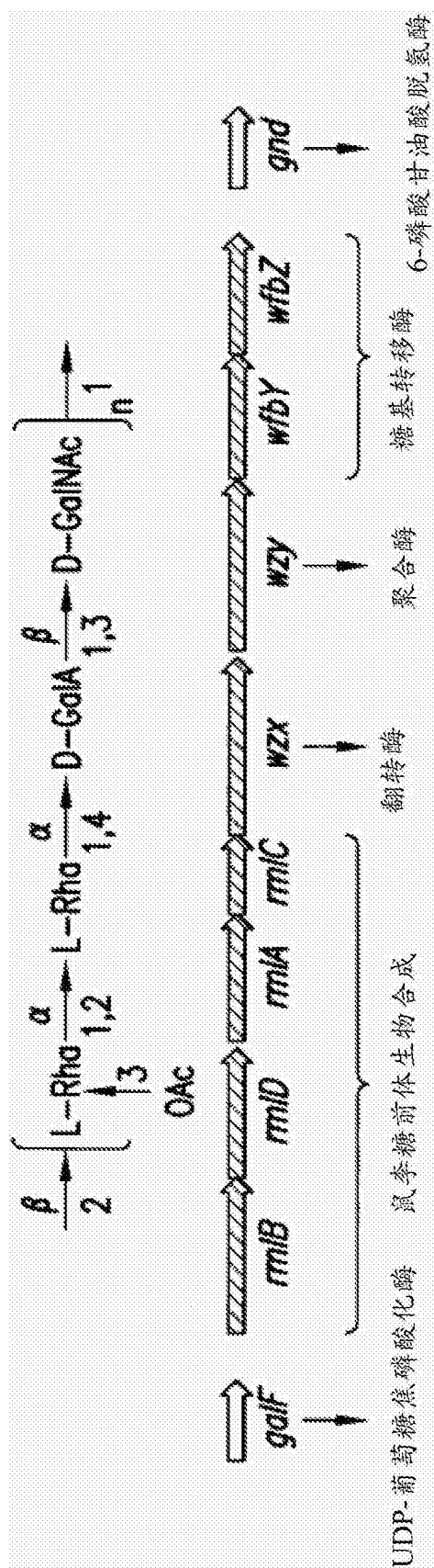


图9

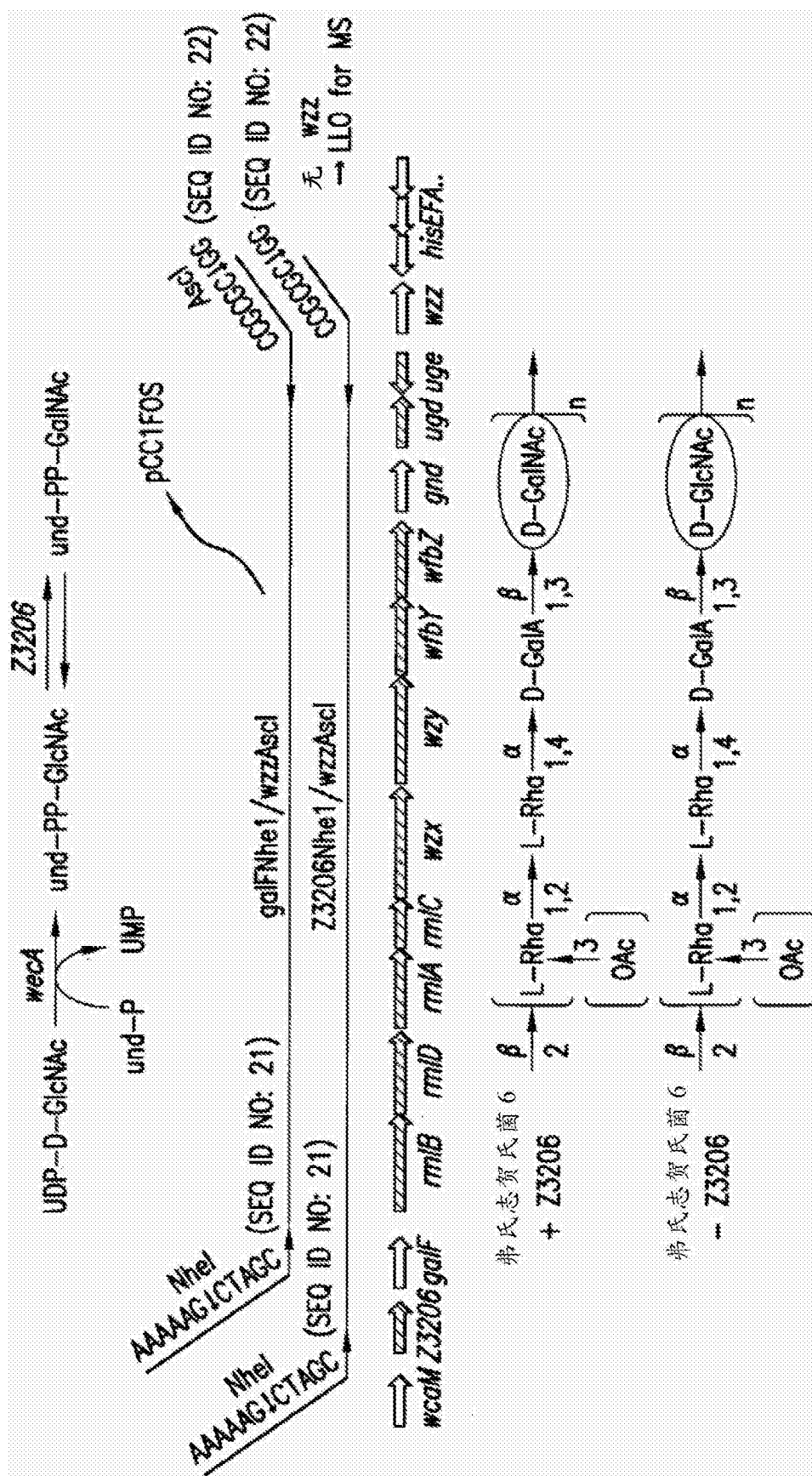


图10

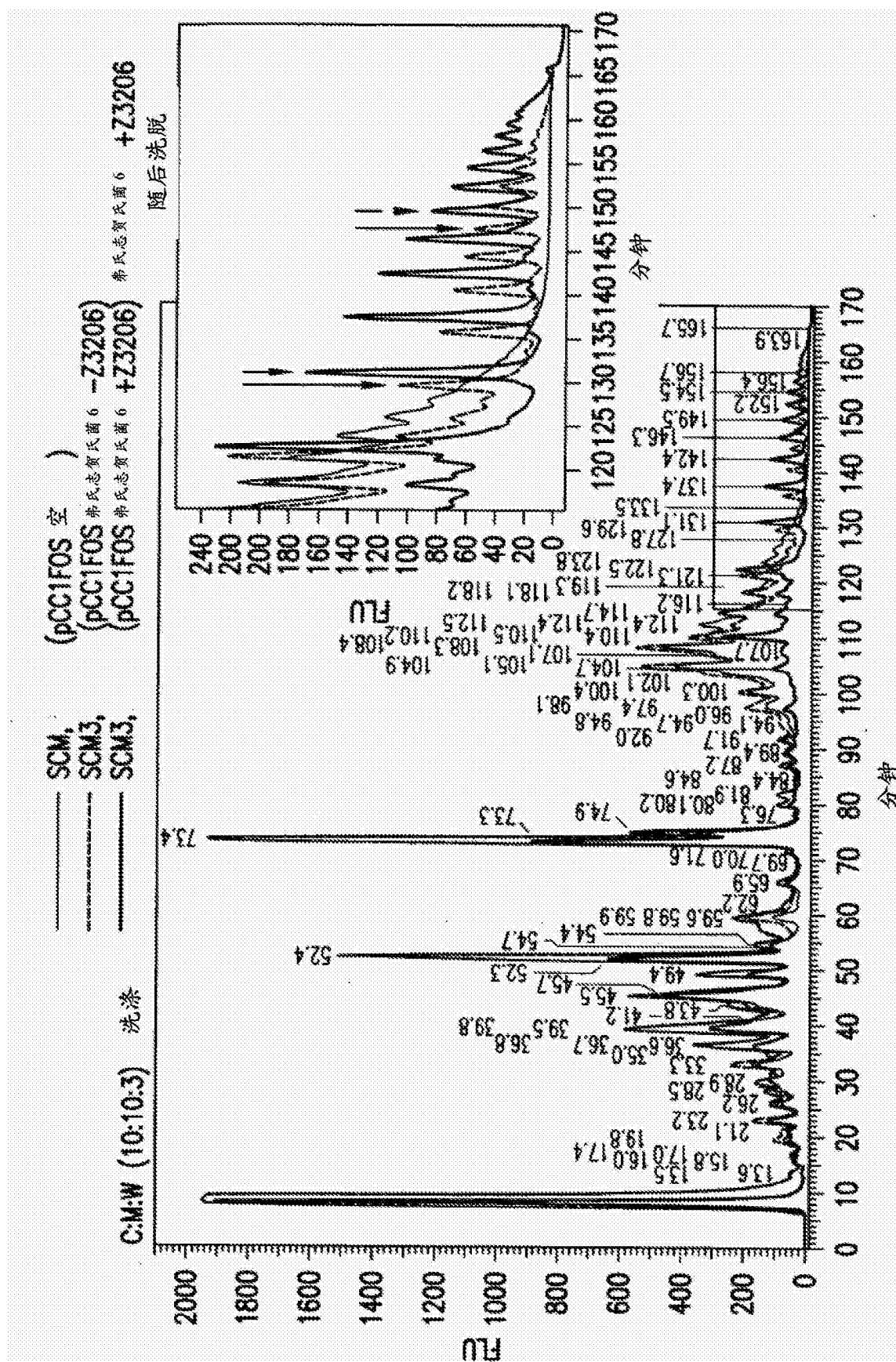


图12

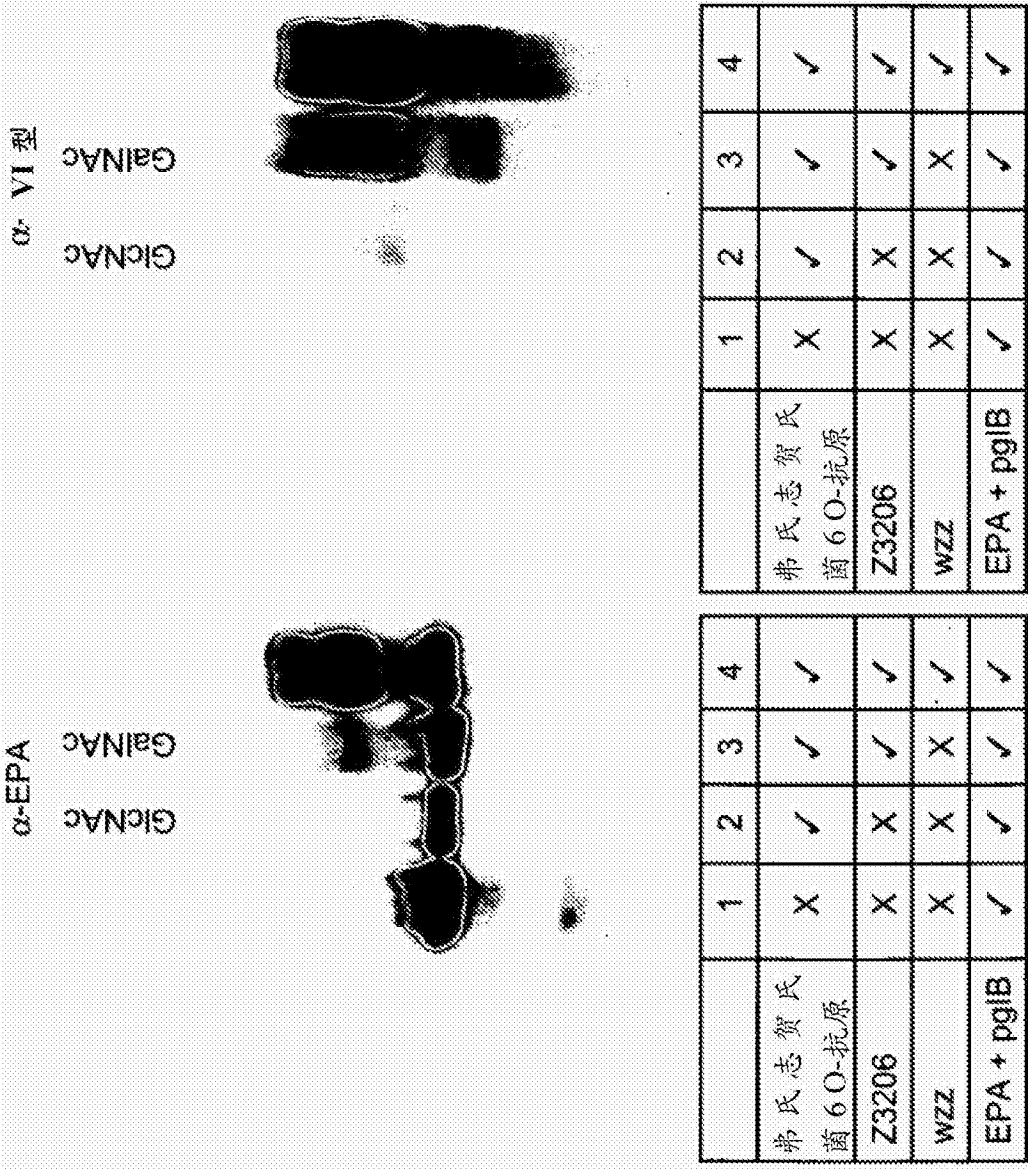


图13